AGGREGATION AND ASSUMPTIONS ON MACRO FUNCTIONS

by

Thomas M. Stoker
Massachusetts Institute of Technology

November 1980
WP 1177-80

ABSTRACT

This paper studies the aggregation structure necessary for several common assumptions on macroeconomic functions to be valid. Sufficient conditions are found for linearity of a macro function, including linear functional form (LFF) and linear probability movement (LPM) structures. Necessary and sufficient conditions are found for omitting a distribution parameter from a macro function. These conditions often provide testable covariance restrictions when the distribution form is known. For the case of unknown distribution form, necessary and sufficient conditions are found for a macro function to depend only on a predictor variable mean, on marginal distributions of jointly distributed variables, and on overall means when the predictor distribution is segmented. A simple model is presented to analyze changing domain of predictor variables. Finally, issues of interpreting and constructing macroeconomic functions are discussed.

KEY WORDS

Aggregation Theory

Linear Functional Form

Linear Probability Movement

Probability Density Decomposition

* Thomas M. Stoker is Assistant Professor, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, 02139. The author wishes to thank Zvi Griliches for early discussion of this topic, and Julio Rotemberg for several helpful discussions.

AGGREGATION AND ASSUMPTIONS ON MACRO FUNCTIONS

1. INTRODUCTION

When formulating macroeconomic models[1] of average data, economists usually have a theory pertaining only to the behavior of a rational individual. If a model based on individual behavior is directly implemented with average data, the researcher implicitly makes a number of strong assumptions about the distributional influences on the average data. For example, suppose that a theory states that consumption is an increasing function of disposable income for each individual in the population. If the researcher then regresses average consumption on average disposable income, he has assumed first that all changes in the disposable income distribution (other than the average) are irrelevant to average consumption, and second that the true relation between average consumption and average disposable income is linear.[2] Failure of either of these assumptions removes any structural or behavioral interpretation of the results of the regression.[3]

In general the movement of an average dependent variable depends on both the true behavioral relationship for individuals and all changing parameters of the predictor variable distribution. In the above example, average consumption will usually depend not only on average income, but also any other changing income distribution parameter, such as the variance, skewness, and (if the elderly save differently than others) the percentage of income accounted for by the elderly. Consequently, correctly modeling an average-dependent variable not only requires knowledge of the individual behavioral relation, but may also require a large amount of detailed information on the underlying distribution movement.

Frequently assumptions are made which allow a macro function to be written in a simplified form. Typically, these assumptions allow all distribution parameters other than predictor means to be omitted from the aggregate formulation. Also, common assumptions allow the basic structural macro model to be written in linear form.

There are several reasons why such simplifying assumptions are either convenient or necessary to use. First, the vast majority of statistical techniques available for estimation and testing of macroeconomic models rely on a linear structural form as their foundation. Second, even if the predictor variable distribution is observed for all time periods under study, it may be computationally infeasible (or too costly) to capture all of the observed detail in the average variable analysis.[4] Alternatively, and usually the case, only partial aspects of the distribution are observed, requiring a simplified aggregate model by necessity.

The purpose of this paper is to establish the structure which underlies several common simplifying assumptions made on macroeconomic relations. This study requires a direct investigation of the interaction between the individual behavioral relation and the movement of the predictor variable distribution through the process of aggregation. Each simplifying assumption generally involves restrictions on the individual relationship between the dependent and independent variables, restrictions on the movement of the independent variable distribution, or a combination of both.

For example, to insure that average consumption depends only on average income, one could assume that consumption is a linear function of income for every individual, with coefficients constant across individuals. Average consumption is then the same linear function of average income. Restrictions of this type are the focus of linear aggregation theory,[5] and allow such simplified aggregate functions for arbitrary movements in the underlying distribution.

Alternatively, there are situations where the distribution movement is restricted, making linearity assumptions as above unnecessarily strong. If movements in the income distribution are completely parameterized by mean income, then average consumption depends only on mean income regardless of the individual consumption-income relationship.[6]

The above discussion points out two types of assumptions which guarantee a particular form of macro function. The existing literature on aggregation in economics[7] essentially provides several sets of sufficient conditions such as those above. Here, through a general formulation, we study the precise structure underlying all such sets of sufficient conditions.

After presenting the notation and formally stating the problem, we begin the exposition by discussing the types of assumptions which guarantee a linear aggregate function. We next turn to the general conditions allowing parameters to be omitted from aggregate functions. Here the first aspect discussed is the conditions allowing a particular parameter to be omitted when the distribution form is known; and illustrated by applying them to distributions of the exponential family. The second aspect discussed is the conditions under which the aggregate function depends only on certain parameters with the distribution form unknown. This problem is technically quite complex, and so characterizing theorems are shown for discrete distributions only, which allow easier interpretation.

These theorems are then applied to three classes of problems, picked for their practical applicability. First is the problem of when the aggregate function depends only on the mean of the predictor variable distribution. This is clearly the most prevalent assumption in empirical studies of aggregate relationships. After applying the general theorems, several examples are used to illustrate the required structure. Next, we refine the general theorems to focus on the linearity of the aggregate function in the predictor variable mean. Finally, the structure required for the use of several predictor variable means is seen to generalize the earlier development only slightly.

The second class of problems arises when averaging is performed over the joint distribution of two or more variables. Here the researcher may not observe the entire joint distribution, but rather only the marginal distribution of each relevant variable. For example, consumption for each individual may depend on both his income and his age, requiring average consumption to depend on the joint distribution of age and income in the population. If data on the marginal distributions of age and of income are all that is available, then a simplified aggregate function is required. We apply the characterization theorems to this problem, and then illustrate the structure with several examples. We close by discussing when the individual component means may be used in place of the full marginal distributions.

The third problem concerns when segmented distribution information can be ignored. Returning to the case where consumption depends only on income, suppose that average income data for urban consumers and for rural consumers are separately observed, as well as the percentage of rural consumers in the population. When will average consumption depend only on overall average income, allowing the segmented distribution detail to be ignored? We analyze this problem for general distributions, and illustrate the solution by applying the characterization theorems applicable to discrete distributions for this case.

Each of these problems is studied assuming that the domain of the predictor variables does not change over time. We continue the exposition by discussing the modifications of our analysis induced by a changing domain, and present a simple model to study these modifications. This model is then applied to the simplest form of domain movement, namely through translation and scaling.

We conclude with a summary of our results and a discussion of their potential uses. Our results not only generalize all previous studies of aggregation in economics, but also indicate problems in interpreting estimated macro relations, plus the correct structures to assume and test in the construction of macroeconomic models.

2. NOTATION

We begin by setting up the basic notation required for our study. For each time period and each individual agent, there is a dependent quantity of interest x determined by an underlying variable A through an individual behavioral function $x = x(A)$. We assume that A captures all relevant differences across agents, so that the function $x = x(A)$ may be assumed identical across individual agents and time periods. In addition, A is assumed to be a scalar variable for simplicity, and we will consider several predictor variables when the need arises.[8]

In each time period, we assume that the quantity A for each agent represents a random drawing from a distribution with density $p(A|\theta)$. θ is a vector of parameters indicating how the distribution changes over time periods. We assume that the number of agents in the economy is large, so that average variables can be associated with their population expectations.[9] Consequently the average of x, denoted $\phi(\theta)$, is written as

$$\phi(\theta) = E(x) = \int x(A)p(A|\theta)dA \tag{2.1}$$

$E(x) = \phi(\theta)$ is the true macro relation connecting average x to the changing distribution parameters θ. In the consumption-income example, x represents consumption, A income and $p(A|\theta)$ the income distribution. θ can have as components mean income, the variance of income, and any other changing distributional parameter such as skewness or kurtosis.

For general forms of the distribution $p(A|\theta)$ and the behavioral function $x(A)$, the form of $\phi(\theta)$ can be arbitrary, depending on all components of the vector θ. Our primary interest in this paper are the conditions under which $\phi(\theta)$ can be simplified, either to a function linear in the components of θ, or by allowing certain components of θ to be omitted from $\phi(\theta)$.

Before discussing linearity of $\phi(\theta)$, three points require mention. The first is the role of time. While not explicitly included in the notation, time is a crucial element in the interpretation of θ. With a given function $x(A)$, the function $\phi(\theta)$ is determined entirely by $p(A|\theta)$ which changes only through the parameter θ. Thus θ could be denoted with a time subscript, which, while unnecessary for the formulae to follow, would remind the reader of this interpretation. Also note that nothing is said of the particular value of A for any agent in any time period. All that is required to determine $\phi(\theta)$ is $p(A|\theta)$, and thus particular agents can interchange positions randomly over time periods.

The second point regards the constancy of the function $x(A)$. While A captures all relevant differences across agents in any one time period, there may be common parameters, such as prices, which change the form of $x(A)$ over time. While these parameters could easily be incorporated in the form of $x(A)$,[10] they are unnecessary for considering the distributional effects on $E(x) = \phi(\theta)$, so we omit them for simplicity.

Finally, note that the formulation (2.1) of $\phi(\theta)$ assumes a constant domain for the variable A,[11] with θ acting only on the density $p(A|\theta)$. This is crucial to the results of Sections 3-4.4, and while some of the required modifications are indicated in Section 5, the general solutions for varying domain are relegated to future research.

3. LINEAR AGGREGATE FUNCTIONS

In this section we consider θ to be a single scalar parameter, and establish sufficient conditions for $E(x) = \phi(\theta) = a + b\theta$, with a and b constant coefficients over time. This presentation is designed to illustrate both assumptions on the form of $x(A)$ and the form of $p(A|\theta)$, which will appear frequently in subsequent sections.

If θ is the mean of A; $\theta = E(A)$, and the behavioral function is linear in A, i.e. $x(A) = a + bA$, then clearly the macro function is linear in θ, via $\phi(\theta) = E(x) = a + bE(A) = a + b\theta$. This structure on $x(A)$ is referred to as a linear functional form (LFF), and guarantees a linear aggregate function for all forms of the density $p(A|\theta)$.[12]

In order to discover other conditions guaranteeing $\phi(\theta) = a + b\theta$, note that a necessary and sufficient condition is that the second derivative of ϕ vanishes.[13] Assuming p to be twice differentiable in θ and that derivatives may be passed under the integral sign, this condition appears as:

$$0 = \frac{\partial^2 \phi}{\partial \theta^2} = \int x(A) \frac{\partial^2 p}{\partial \theta^2} \, dA \tag{3.1}$$

A second set of conditions guaranteeing the linearity of ϕ are when $\frac{\partial^2 p}{\partial \theta^2} = 0$ for all A. This implies that $p(A|\theta)$ can be written as:

$$p(A|\theta) = p_1(A) + \theta p_2(A) \tag{3.2}$$

where p_1 and p_2 do not depend on θ. Since $\int p(A|\theta)dA = 1$ for all θ; $\int p_1(A)dA + \theta \int p_2(A)dA = 1$ for all θ, which implies $\int p_1(A)dA = 1$ and $\int p_2(A)dA = 0$. Consequently $p_1(A)$ represents a base density,[14] and $\theta p_2(A)$ a distribution shift, occurring linearly with respect to θ. The structure (3.2) is referred to as linear probability movement (LPM), and represents movement of $p(A|\theta)$ over time by simple extrapolation with respect to θ. $\phi(\theta)$ in this case is:

$$\begin{aligned}\phi(\theta) &= \int x(A)p_1(A)dA + \theta \int x(A)p_2(A)dA \\ &= a + \theta b\end{aligned}$$

with a, b constant over time. This form of ϕ is guaranteed for all forms of

$x = x(A)$, and thus LPM represents the "opposite" of LFF, which insures linearity for all distribution forms.

Any combination $x(A)$, $p(A|\theta)$ obeying condition (3.1) will produce a linear macro function ϕ. For example. suppose that $p_3(A)$ is any function orthogonal to $x(A)$; i.e. $\int x(A)p_3(A)dA = 0$. Then

$$p(A|\theta) = p_1(\theta) + \theta\ p_2(A) + d(\theta)p_3(A) \tag{3.3}$$

is a density[15] whose movement insures a linear macro function. Viewing (3.3) from the standpoint of a fixed function $p_3(A)$, a linear macro function is guaranteed for any behavioral function $x(A)$ orthogonal to $p_3(A)$. This interdependence of distribution and functional form is typical of the problems studied here, and reappears in future sections.

Notice in addition that if θ is a K-vector of several parameters, a multivariate version of LPM can be defined, with $p(A|\theta)$ moving over time by a linear combination of K shift functions. Similarly if θ represents the means of K underlying variables, a corresponding form of LFF is defined through x being a linear function of these underlying variables. Either LPM or LFF in this form guarantees an aggregate function $\phi(\theta)$ linear in the components of θ.

We now turn to a discussion of omitting components of θ from $E(x) = \phi(\theta)$.

4. OMITTING PARAMETERS

We begin our discussion of omitting parameters from $E(x) = \phi(\theta)$ with a general development and then turn to the specific problems mentioned in the introduction. θ is a K vector, here partitioned as $\theta = (\theta_0, \theta_1)$, where θ_0 is a K_0 vector and θ_1 a $K_1 > 0$ vector, $K_0 + K_1 = K$. Our interest is in the conditions under which $E(x)$ is a function of θ_0 only; i.e. $\phi(\theta)$ appears as:

$$\phi(\theta) = \int x(A)p(A|\theta)dA = \phi_0(\theta_0)$$

where ϕ_0 does not depend on θ_1.

This property is clearly guaranteed if $p(A|\theta)$ depends only on θ_0, or if there exists a functional relation $\theta_1 = g(\theta_0)$ between θ_1 and θ_0 for all time periods. To remove these uninteresting cases from our analysis, we adopt:

Assumption 1: $\Theta = \{\theta\varepsilon\ R^K \mid p(A|\theta)$ is a density$\}$ has a nonempty interior in $\mathbb{R}^K$. Also, there exists $\theta^1 = (\theta_0^1,\theta_1^1)$ and $\theta^2 = (\theta_0^2,\theta_1^2)$, $\theta^1,\theta^2 \varepsilon\ \Theta$ such that $\theta_1^1 \neq \theta_1^2$ and $p(A|\theta^1) \neq p(A|\theta^2)$ for all A.

For our discussion of general forms $x(A)$ and densities $p(A|\theta)$,[16] we must also adopt:

Assumption 2: $\phi(\theta)$ and $p(A|\theta)$ are differentiable in the components of θ_1 for all A and $\theta\varepsilon\ \Theta$ and differentiation may be performed under the integral defining $\phi(\theta)$.

Now, under assumptions 1 and 2, θ_1 can be omitted from $E(x) = \phi(\theta)$ if and only if the gradient $\nabla_{\theta_1}\phi$ vanishes for all $\theta\varepsilon\ \Theta$. This gradient can be expressed in the following two equivalent ways:

$$0 = \nabla_{\theta_1}\phi = \int x(A)\nabla_{\theta_1} p(A|\theta)dA \tag{4.1a}$$

$$= \mathrm{Cov}(x(A),\nabla_{\theta_1} \ln p) \tag{4.1b}$$

where the latter equality follows from $E(\nabla_{\theta_1} \ln p) = 0$. Equation (4.1a) states that $x(A)$ must be orthogonal to each component of $\nabla_{\theta_1} p$ via the integral inner product. This version is useful in our later analysis, where the density p is orthogonally decomposed and θ_0 and θ_1 related to the decomposition. Equation (4.1b) states that $x(A)$ and each component of the score vector $\nabla_{\theta_1} \ln p$ are uncorrelated. This form is useful when the form of the density p is known, as illustrated by the following examples:

Example 1:

Suppose that p is a member of the exponential family in its natural parameterization:[17]

$$p = C(\pi)h(A)\exp\left(\sum_{i=1}^{K} \pi_i v_i(A)\right)$$

where $\pi = (\pi_1,\ldots,\pi_K)$ is a vector of parameters and $v = (v_1(A),\ldots,v_K(A))$ is a vector function of A. Let $\theta = \pi$ with $\theta_0 = (\pi_1,\ldots,\pi_{K-1})$ and $\theta_1 = \pi_K$. Then $E(x) = \phi(\pi)$ does not depend on π_K if and only if

$$\begin{aligned} 0 &= \mathrm{Cov}\ (x(A),\ \frac{\partial \ln p}{\partial \pi_K}) \\ &= \mathrm{Cov}\ (x(A),\ \frac{\partial \ln C(\pi)}{\partial \pi_K} + v_K(A)) \\ &= \mathrm{Cov}\ (x(A),\ v_K(A)) \end{aligned}$$

i.e. $x(A)$ and $v_K(A)$ are uncorrelated for all $\pi = \theta \varepsilon \Theta$.

Example 2:

Suppose that p is a member of the exponential family as above. From a practical point of view, it is more interesting to inquire when $E(x)$ is determined by the means of $v_i(A)$, $i = 1, \ldots, K-1$, omitting the mean of $v_K(A)$. Denoting as $\mu = (\mu_1,\ldots,\mu_K)'$ the vector of means $(\mu_i = E(v_i(A)),\ i=1,\ldots,K)$ we reparameterize the density p with respect to μ, and set $\theta = \mu$, with $\theta_0 = (\mu_1,\ldots,\mu_{K-1})$, $\theta_1 = \mu_K$. It can be shown that the gradient of ϕ with respect to the full vector θ is[18]

$$\nabla_\theta \phi = E((v-\mu)(v-\mu)')E((x-\phi(\theta))(v-\mu))$$

the vector of "slope regression coefficients." Thus $\theta_1 = \mu_K$ can be omitted from $E(x) = \phi(\theta)$ if and only if its associated regression coefficient is zero.

Example 3:

Suppose that p is a normal density with mean μ and variance σ^2. Our interest is in when σ^2 can be omitted from E(x), so let $\theta = (\mu,\sigma^2)$, $\theta_0 = \mu$, $\theta_1 = \sigma^2$. Since $\ln p = -\ln 2\pi - \frac{1}{2}\ln\sigma^2 - \frac{1}{2}\frac{(A-\mu)^2}{\sigma^2}$

$$\mathrm{Cov}\left(x, \frac{\partial \ln p}{\partial\sigma^2}\right) = \frac{1}{2\sigma^4}\,\mathrm{Cov}\ (x,(A-\mu)^2)$$

Thus E(x) depends only on $\theta_0 = \mu$ if and only if x(A) is uncorrelated with the squared derivation $(A-\mu)^2$ for all μ,σ^2. Unfortunately this is all that can be said in general. If x(A) is linear in A, then clearly this covariance vanishes: let x(A) = a + bA, then

$$\mathrm{Cov}((a+bA),(A-\mu)^2) = bE((A-\mu)^3) = 0$$

for all μ,σ^2. Also, as can be easily verified, if x(A) is quadratic in A, then it is impossible to omit σ^2 from E(x).

In each of these examples, equation (4.1b) was used to obtain conditions requiring a specific covariance to vanish for all $\theta\varepsilon\Theta$. With cross-section data for a single time period, these covariances can be estimated consistently, and thus the condition that they vanish tested. Consequently, when the density form p is known or assumed, the conditions for omitting parameters can be tested, with rejection of such conditions indicating that the respective distribution parameters must be included.[19]

While (4.1b) is a valid condition for all problems of omitting parameters, it is less useful than (4.1a) for analyzing assumptions that $E(x) = \phi(\theta)$ depends only on certain distribution aspects (such as independent variable means) for all possible distributions. For this case, the orthogonality indicated by (4.1a) provides a useful focus. The analysis of this problem for discrete densities follows.

4.1 Parameter Omission with Discrete Distributions

This problem is best addressed by considering the density $p(A|\theta)$ and the function $x(A)$ as vectors. Suppose that there are N possible types of agents, indexed by $i = 1, \ldots, N$, with the i^{th} type occurring in the population with probability (or proportion) $p_i(\theta)$. The vector of probabilities is denoted $\underline{p}(\theta) = (p_1(\theta),\ldots,p_N(\theta))'$ and lies in the unit simplex $T = \left\{ \underline{p} \varepsilon R^N \,\middle|\, \underline{i}'\underline{p} = 1, \underline{p} \geq 0 \right\}$ where $\underline{i} = (1,\ldots,1)'$.

Any random variable taking values for the N possible consumer types can be represented by an N vector. Denote by $\underline{A} = (A_1,\ldots,A_N)'$ the independent quantity of interest, and by $\underline{x} = (x(A_1),\ldots,x(A_N))'$ the dependent quantity.[20]

Expectations appear simply as inner products:

$$E(A) = \underline{A}'\underline{p}(\theta)$$

$$\text{and} \quad E(x) = \underline{x}'\underline{p}(\theta) = \phi(\theta)$$

Our general strategy in studying parameter omission will be to express the movement of $\underline{p}(\theta)$ in T in terms of N-1 orthogonal directions. These directions are then related to the parameter subvectors θ_0 and θ_1. We begin by noting that

$$\underline{p}(\theta) = \sum_{i=1}^{N} p_i(\theta)\underline{e}_i \tag{4.2}$$

where $e_i = (0,\ldots,1\ldots,0)'$ is the i^{th} unit vector of R^N. $\left\{\underline{e}_1,\ldots,\underline{e}_N\right\}$ represents an orthogonal basis of R^N,[21] with $p_i(\theta)$ the coefficient of $\underline{e}_i$. In order to guarantee that $\underline{i}'\underline{p}(\theta) = 1$, we choose N-1 vectors $\underline{\delta}_i$, $i = 2, \ldots, N$, such that $\left\{\underline{i}, \underline{\delta}_2, \ldots, \underline{\delta}_N\right\}$ is an orthogonal basis of R^N. (4.2) above can be written uniquely as:

$$\underline{p}(\theta) = \frac{1}{N}\underline{i} + \sum_{i=2}^{N} d_i(\theta)\, \underline{\delta}_i \tag{4.3}$$

where $d_i(\theta)$, $i = 2, \ldots, N$ are the transformed coefficients from (4.2). Since $\underline{i}'\underline{\delta}_i = 0$, $i = 2, \ldots, N$, the second term of (4.3) represents the movement of $\underline{p}(\theta)$ in the unit simplex T, with $\underline{\delta}_2, \ldots, \underline{\delta}_N$ the orthogonal directions of movement and $d_i(\theta)$ gauging the amount of movement in direction i.[22]

Recall that $\theta = (\theta_0, \theta_1)'$ and that our interest is in determining when $E(x) = \phi(\theta) = \phi_0(\theta_0)$. For this purpose the decomposition (4.3) must be altered to be minimal in θ_1 in the following sense.

Theorem 1: Given Assumption 1, there exists a decomposition of $\underline{p}$:

$$\underline{p}(\theta) = \frac{1}{N}\underline{i} + \sum_{i=1}^{N_1} f_i(\theta_0)\underline{\delta}_i + \sum_{j=1}^{N_2} g_j(\theta_0, \theta_1)\, \underline{\xi}_j$$

where $\underline{i}$, $\underline{\delta}_i$, $i = 1, \ldots N_1$, $\underline{\xi}_j$, $j = 1, \ldots N_2$ are orthogonal vectors and $N_1 + N_2 \leq N-1$, such that if

$\Sigma\alpha_j g_j(\theta_0, \theta_1) = f(\theta_0)$, then $\alpha_j = 0$ for $j = 1, \ldots, N_2$.

Such a decomposition is called minimal in θ_1. Moreover, given any other decomposition which is minimal in θ_1, say

$$\underline{p}(\theta) = \frac{1}{N}\underline{i} + \sum_{i=1}^{N_1^*} f_i^*(\theta_0)\underline{\delta}_i^* + \sum_{j=1}^{N_2^*} g_j^*(\theta_0, \theta_1)\underline{\xi}_j^*$$

then $N_2^* = N_2$, and [23]

$$\text{SPAN}\left\{\underline{\xi}_1, \ldots, \underline{\xi}_{N_2}\right\} = \text{SPAN}\left\{\underline{\xi}_1^*, \ldots, \underline{\xi}_{N_2}^*\right\}$$

Proof: See the Appendix.

The implication of this theorem is that the action of θ_1 on $\underline{p}(\theta)$ can be isolated to a unique subspace of directions $S_{\theta_1} = \text{SPAN} \left\{ \xi_1, \ldots, \xi_{N_2} \right\}$. The result allowing use of this decomposition is:

Theorem 2: Let $\underline{x} \varepsilon R^N$ represent a random variable x. $E(x) = \phi_0(\theta_0)$ if and only if $x \varepsilon S_{\theta_1}^{\perp}$, where $S_{\theta_1}^{\perp}$ in the subspace orthogonal to S_{θ_1} in R^N.

Proof: See the Appendix.

When discussing several parameterizations of a given density, another useful property is given by:

Theorem 3: Suppose that θ_1 and θ_1' are K_1 vectors of parameters, and there exists a function $\theta_1' = g(\theta_1)$ which is 1 - 1. Then if $\underline{p}(\theta)$ is reparameterized to $\underline{p}'(\theta')$, with $\theta' = (\theta_0, \theta_1')$, then $S_{\theta_1} = S_{\theta_1}'$.

Proof: See the Appendix.

Thus, a 1-1 reparameterization of p holding θ_0 constant does not alter the subspace of contaminated directions S_{θ_1}.

This machinery allows us to analyze the substantive problems raised in the Introduction.

4.2 Aggregate Functions Depending only on Predictor Variable Means

In our basic setup, A is a scalar variable whose mean we denote by μ. Recall that

$$\mu = \underline{A}'\underline{p}(\theta) \tag{4.4}$$

where $\underline{A}$ is the vector representing A and $\underline{p}(\theta)$ is the vector of probabilities. For this problem we assume that $\theta = (\mu, \theta_1)$,[24] and consider the conditions under which $E(x) = \phi_0(\mu)$.

We study this problem by finding a unique direction vector $\tilde{\underline{A}}$ associated with μ for all density vectors obeying (4.4). $\underline{\tilde{A}}$ is found by orthogonally decomposing $\underline{A}$ as:

$$\underline{A} = \bar{A}\underline{i} + (\underline{A} - \bar{A}\underline{i})$$
$$= \bar{A}\underline{i} + \underline{\tilde{A}}$$

with $\bar{A} = \sum_{i=1}^{N} A_i/N$, the (unweighted) average of possible values of A. Now, for any density vector $\underline{p}(\mu,\theta_1)$, choose N-2 vectors $\underline{\xi}_3,\ldots,\underline{\xi}_N$ such that $\{\underline{i}, \underline{\tilde{A}}, \underline{\xi}_3, \ldots, \underline{\xi}_N\}$ is an orthogonal basis of $\mathbf{R}^N$, and write $\underline{p}(\mu,\theta_1)$ as:

$$\underline{p}(\mu,\theta_1) = \frac{1}{N}\underline{i} + D^*(\mu,\theta_1)\underline{\tilde{A}} + \sum_{i=3}^{N} d_i(\mu,\theta_1)\,\underline{\xi}_i \tag{4.5}$$

where D^* and d_i, i = 3, ..., N represent the relevant coefficients. That $\underline{\tilde{A}}$ is a unique direction associated with μ is shown via (4.4) and (4.5) as:

$$\mu = \underline{A}'\underline{p}(\mu,\theta_1) = \frac{1}{N}\bar{A}\underline{i}'\underline{i} + D^*(\mu,\theta_1)\,\underline{\tilde{A}}'\underline{\tilde{A}}$$
$$= \bar{A} + D^*(\mu,\theta_1)\,\underline{\tilde{A}}'\underline{\tilde{A}}$$

so

$$D^*(\mu,\theta_1) = (\mu-\bar{A})/\underline{\tilde{A}}'\underline{\tilde{A}} = D(\mu)$$

Thus $D^*(\mu,\theta_1) = D(\mu)$ depends only on μ in a linear way. Note that the form of $D(\mu)$ is the same for all density vectors obeying (4.4). Finally, from Theorem 1, write (4.5) in minimal form as:

$$\underline{p}(\mu,\theta_1) = \frac{1}{N}\underline{i} + D(\mu)\underline{\tilde{A}} + \sum_{i=2}^{N_1} f_i(\mu)\,\underline{\delta}_i + \sum_{j=1}^{N_2} g_j(\mu,\theta_1)\underline{\xi}_j \tag{4.6}$$

where $N_1 + N_2 \le N - 1$.

Before presenting the main result of this section, recall that the function x(A) is represented by an N vector $\underline{x}$. $\underline{x}$ can be written uniquely as:

$$\underline{x} = \bar{x}\,\underline{i} + b\underline{\tilde{A}} + \underline{\varepsilon} \tag{4.7}$$

where $\bar{x} = \sum_{i=1}^{N} x(A_i)/N$ and $\underline{\varepsilon}'\underline{i} = \underline{\varepsilon}'\underline{\tilde{A}} = 0.$[25]

The main result of this section is:

Theorem 4: Given Assumption 1, if $\underline{p}(\mu,\theta_1)$ is in the minimal form (4.6) and $\underline{x}$ is of the form (4.7), then $E(x) = \underline{x}'\underline{p}(\mu,\theta_1) = \phi_0(\mu)$ if and only if $\underline{\varepsilon}'\underline{\xi}_j = 0$, $j = 1, \ldots, N_2$

Corollary 5:

(a) If $E(x) = \phi_0(\mu)$ for all distributions obeying $E(A) = \mu$, then $x(A) = a + bA$, where a and b are constants, and $\phi_0(\mu) = a + b\mu$.

(b) Any density vector such that $E(x)$ depends only on μ for all functions $x = x(A)$ must depend only on μ, and no other varying parameters.

Proof: Theorem 4 follows directly from Theorem 2. For Corollary 5(a), pick N-2 vectors $\underline{\xi}_j$, $j = 3, \ldots, N$ such that $\left\{ \underline{i}, \underline{\tilde{A}}, \underline{\xi}_3, \ldots, \underline{\xi}_N \right\}$ is an orthogonal basis of R^N, and consider the density,[26]

$$\underline{p}(\mu,d_3,\ldots,d_N) = \frac{1}{N}\,\underline{i} + D(\mu)\underline{\tilde{A}} + \sum_{i=3}^{N} d_i\underline{\xi}_i$$

From Theorem 4, $E(x) = \phi_0(\mu)$ iff

$$\underline{x} = \bar{x}\ \underline{i} + b\underline{\tilde{A}} = (\bar{x} - \bar{A})\ \underline{i} + b\underline{A}$$

This is equivalent to the functional form $x(A) = a + bA$, where $a = \bar{x} - \bar{A}$ and b are constants. For (b) suppose that $\underline{p}(\mu,\theta_1)$ represents a density such that $\underline{p}$ depends nontrivially on θ_1 and $E(x)$ depends only on μ for all functions $x = x(A)$. Write $\underline{p}(\mu,\theta_1)$ in minimal form as:

$$\underline{p}(\mu,\theta_1) = \frac{1}{N}\,\underline{i} + D(\mu)\ \underline{\tilde{A}} + \sum_{i=2}^{N_1} f_i(\mu)\ \underline{\delta}_i + \sum_{j=1}^{N_2} g_j(\mu,\theta_1)\ \underline{\xi}_i$$

If we choose $\underline{x} = \underline{\xi}_1$, then

$$E(x) = \underline{x}'\underline{p}(\mu,\theta_1) = g_1(\mu,\theta_1)\ \underline{\xi}_1'\underline{\xi}_1$$

which is a contradiction, so we must have

$$\underline{p}(\mu,\theta_1) = \underline{p}(\mu), \text{ depending only on } \mu.$$

Q.E.D.

Corollary 5(a) is a refinement of the Fundamental Theorem of Exact Aggregation of Lau (1980), which states that for an aggregate function to depend only on μ for unrestricted distribution movement, the behavioral function $x(A)$ must be linear in A (i.e. LFF holds). Corollary 5(b) states that if a distribution moves with respect to a parameter other than μ, then $E(x)$ is a function of μ only if $x(A)$ is restricted according to Theorem 4.

We now illustrate the density decomposition and the above theorem with several examples.

Example 4 (N=2):

Assume that A is zero with probability $1-P$ and one with probability P, so $\underline{p} = (1-P,P)'$, $\underline{A} = (0,1)'$, and $\mu = E(A) = P$. Here $\bar{A} = \frac{1}{2}$; $\underline{A} - \bar{A}\underline{i} = \underline{\tilde{A}} = (-\frac{1}{2},\frac{1}{2})'$ and the decomposition (4.6) appears as

$$\underline{p} = \frac{1}{2}\underline{i} + D(P)\underline{\tilde{A}}$$

with $D(P) = (P-\bar{A})/\underline{\tilde{A}}'\underline{\tilde{A}} = 2P-1$. The direction vector $\underline{\tilde{A}}$ indicates that $\underline{p}$ changes only by increasing P and decreasing $1-P$ simultaneously (or vice versa) in equal amounts. If $x(0) = x_0$ and $x(1) = x_1$, then $\underline{x} = (x_0,x_1)'$, and $E(x)$ is a function of P if $x_0 \neq x_1$, i.e. $x(A)$ is not a constant function, with $\underline{x}$ collinear with $\underline{i}$. Notice in addition that if A^* was the relevant predictor variable where $\underline{A}^* = (a,b)'$, $a \neq b$, then $\mu = E(A^*) = a + (b-a)P$, and the only changes in the above decomposition would be $\underline{\tilde{A}}^* = (b-a)\underline{\tilde{A}}$ and $D^*(\mu) = (2\mu-a-b)/(b-a)^2$, with conclusions regarding $E(x)$ unchanged.

Example 5 (N=3):

Let $\underline{A} = (0,1,2)'$. If $\underline{p} = (1-P-Q,P,Q)'$ then $\mu = E(A) = P + 2Q$. Parameterizing $\underline{p}$ in terms of $\theta = (\mu,Q)$ gives $\underline{p}(\mu,Q) = (1-\mu+Q,\mu-2Q,Q)'$. Now $\bar{A} = 1$, $\underline{\tilde{A}} = (-1,0,1)'$, and the decomposition (4.6) of $\underline{p}(\mu,Q)$ is

$$\underline{p}(\mu,Q) = \frac{1}{3}\underline{i} + D(\mu)\underline{\tilde{A}} + g(\mu,Q)\underline{\xi}$$

where $D(\mu) = (\mu-\bar{A})/\underline{\tilde{A}}'\underline{\tilde{A}} = (\mu-1)/2$, $g(\mu,Q) = Q + \frac{1}{6} - \frac{\mu}{2}$ and $\underline{\xi} = (1,-2,1)'$. Here the admissible parameter space is:

$$\Theta = \left\{(\mu,Q)\,|\,\mu,Q \in R;\quad 0 \leq Q \leq 1;\ 2Q \leq \mu \leq 1 + 2Q\right\}$$

$\underline{\tilde{A}}$ corresponds to a precentage shift in probability from the first class to the third, and ξ to a shift from the second class to the first and third (in equal amounts). Now, suppose that a function $x = x(A)$ is represented by a vector $\underline{x}$, which in turn can be written as in (4.7):

$$\underline{x} = c\underline{i} + b\underline{\tilde{A}} + e\underline{\xi}$$

$E(x) = \underline{x}'\underline{p}(\mu,Q)$ is therefore:

$$E(x) = \frac{1}{3}c\underline{i}'\underline{i} + bD(\mu)\underline{\tilde{A}}'\underline{\tilde{A}} + eg(\mu,Q)\ \underline{\xi}'\underline{\xi}$$

$$= c + b(\mu-1) + e(6Q+1-3\mu)$$

Clearly $E(x)$ is independent of Q iff $e = 0$, in which case $E(x) = (c-b) + b\mu$ and $x(A) = (c-b) + bA$. Any nonlinear function $x = x(A)$ will require $E(x)$ to depend on Q. Also notice that if the movement in $\underline{p}$ is such that Q is constant, then $E(x)$ is linear in μ for all possible forms of the function $x(A)$.

To illustrate this example graphically, it is more convenient to use a different parameterization of $\underline{p}$, namely

$$\underline{p}(\mu,g) = \frac{1}{3}\,\underline{i} + D(\mu)\,\underline{\tilde{A}}' + g\underline{\xi}$$

where $g = Q + \frac{1}{6} - \frac{\mu}{2}$ is now the relevant parametric aspect. The admissible parameter space is now $\Theta = \left\{(\mu,g)\,|\,\mu,g\varepsilon\ R; -\frac{1}{3} \leq g \leq \frac{1}{6};\ \frac{1}{3} - 2g \leq \mu \leq \frac{5}{3} + 2g\right\}$ The space of possible distributions T is presented in Figure 1. Here the vector $\frac{1}{3}\,\underline{i}$, as well as the directions $\underline{\tilde{A}}$ and $\underline{\xi}$ are noted. If g is fixed and $\mu = 1$, the density vector $\underline{p}$ lies along segment 12, say at 3. If over time, the density vector moves in a nonlinear way, say along curve 34, then $E(x)$ is a function only of μ if $\underline{x}$ is orthogonal to $\underline{\xi}$ (i.e. $x(A)$ is linear in A) so that the distance traveled along direction $\underline{\tilde{A}}$ (i.e. $D(\mu)$) solely determines $E(x)$. If g is constant and μ varies, the density vector moves along the linear segment 56. In this case, $E(x)$ is a linear function of μ.

If Q instead of g is held constant, the density vector will move over time along a linear segment parallel to $\underline{\eta} = \underline{\tilde{A}} - \underline{\xi} = (-2,2,0)$, and $E(x)$ is a (different) linear function of μ. Each of these cases corresponds to LPM in μ, as does movement along any linear segment (except parallel to 12, where μ is constant), giving $E(x)$ as a linear function of μ.

In this example the role of functional form restrictions (LFF) and distribution movement restrictions (LPM) were illustrated as to how they could insure that $E(x)$ is a (linear) function only of μ. The next example considers orthogonality restrictions other than LFF or LPM.

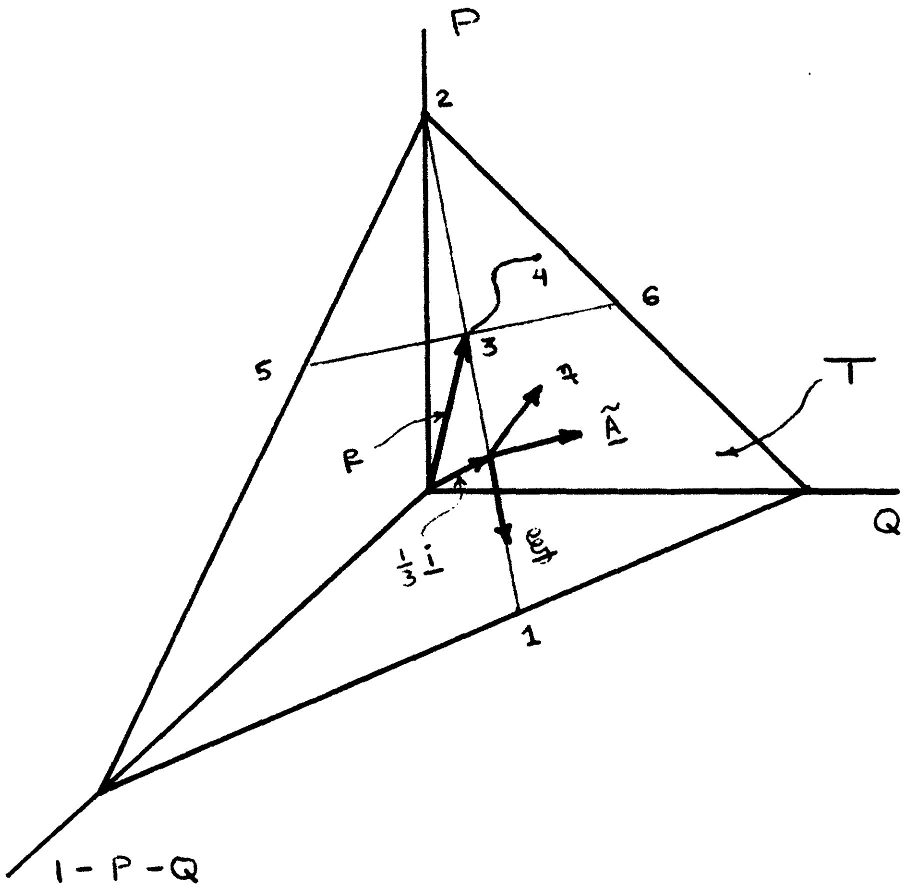

FIGURE 1: Orthogonal Decomposition when there are Three Types of Individuals (N=3).

Example 6 (N=4):

Let $\underline{A} = (2,4,6,8)'$. If $\underline{p} = (1-P-Q-R,P,Q,R)'$ then $\mu = E(A) = 2 + 2P + 4Q + 6R$. Parameterizing $\underline{p}$ in terms of $\theta = (\mu,Q,R)$ gives:

$$\underline{p}(\mu,Q,R) = \begin{pmatrix} 2 - \frac{\mu}{2} + Q + 2R \\ -1 + \frac{\mu}{2} - 2Q - 3R \\ Q \\ R \end{pmatrix}$$

Now $\bar{A} = 5$, $\underline{\tilde{A}} = (-3,-1,1,3)'$ and the decomposition (4.6) appears as:

$$\underline{p}(\mu,Q,R) = \frac{1}{4}\,\underline{i} + D(\mu)\underline{\tilde{A}} + g_1(\mu,Q,R)\underline{\delta}_1 + g_2(\mu,Q,R)\underline{\delta}_2 \qquad (4.8)$$

where $D(\mu) = (\mu-\bar{A})/\underline{\tilde{A}}'\underline{\tilde{A}} = \frac{\mu}{20} - \frac{1}{4}$, and

$$g_1(\mu,Q,R)\underline{\delta}_1 + g_2(\mu,Q,R)\underline{\delta}_2 = \begin{pmatrix} 1 - \frac{7}{20}\mu + Q - 2R \\ -\frac{3}{2} + \frac{11}{20}\mu - 2Q - 3R \\ Q - \frac{\mu}{20} \\ \frac{1}{2} + R - \frac{3\mu}{20} \end{pmatrix}$$

with g_1, g_2, $\underline{\delta}_1$ and $\underline{\delta}_2$ to be specified later.

Now, if $x = x(A)$, then $E(x)$ is a function of μ if and only if $x(A)$ is linear in A, with the vector representation $\underline{x}$ collinear with $\underline{i}$ and $\underline{\tilde{A}}$.

In general if there are N types of consumers, for a given function $x(A)$ there will be N-3 directions in which the density vector can move leaving $E(x)$ unchanged (i.e. the N-3 directions orthogonal to $\underline{\varepsilon}$, $\underline{\tilde{A}}$ and $\underline{i}$ of (4.7)). Also, the N-3 directions are different for different nonlinear functions. We illustrate this in the current example (N=4, N-3=1) by considering two formulations of x(A).

Suppose first that

$$x(A) = b_1 A, \ A \leq 4$$

$$= b_2 A, \ A > 4$$

The reader can verify that $\bar{x} = \frac{3}{2} b_1 + \frac{7}{2} b_2$, and

$$\underline{x} = \bar{x}\underline{i} + \left[b_2 - \frac{b_1 - b_2}{2} \right] \underline{\tilde{A}} + (b_1 - b_2)\,\underline{\delta}_1$$

where $\underline{\delta}_1 = (-1,2,-1,0)'$. E(x) is constant if the density moves in the direction of $\underline{\delta}_2 = (2,-1,-4,3)'$, which is orthogonal to $\underline{i}, \underline{\tilde{A}}, \underline{\delta}_1$. If (4.8) is written with $\underline{\delta}_1$ and $\underline{\delta}_2$ above, then we have:

$$g_1(\mu,Q,R) = -\frac{2}{3} + \frac{1}{4}\mu - Q - \frac{4}{3}R$$

and $$g_2(\mu,Q,R) = \frac{1}{6} - \frac{\mu}{20} + \frac{1}{3}R$$

From the form of g_1, we see that in general E(x) will depend on μ, Q and R. Also from g_1, we see that if the distribution movement is restricted such that $-\Delta Q = \frac{4}{3}\Delta R$,[27] then E(x) is linear in μ.

As mentioned above, $\underline{\delta}_1$ and $\underline{\delta}_2$ with the above property vary with changing functional form. If $x(A) = c_1 A + c_2 A^2$, then $\bar{x} = 5c_1 + 30c_2$ and

$$\underline{x} = \bar{x}\underline{i} + (c_1 + 10c_2)\underline{\tilde{A}} + 4c_2\underline{\delta}_1$$

where $\underline{\delta}_1 = (1,-1,-1,1)'$. Orthogonal to $\underline{i}$, $\underline{\tilde{A}}$ and $\underline{\delta}_1$ is $\underline{\delta}_2 = (-1,3,-3,1)'$, and so any distribution movement in the direction of $\underline{\delta}_2$ leaves E(x) unchanged. If (4.8) is written with $\underline{\delta}_1$, $\underline{\delta}_2$ above, we get:

$$g_1(\mu,Q,R) = \frac{3}{4} - \frac{\mu}{4} + \frac{Q}{2} + \frac{3}{2}R$$

$$g_2(\mu,Q,R) = -\frac{1}{4} - \frac{1}{2}Q - \frac{1}{2}R + \frac{\mu}{10}$$

The same sort of reasoning can be applied as before, in noting that if density movement is restricted by $\Delta Q = -3\Delta R$, then $E(x)$ for this case is a linear function of μ.

Having illustrated the density decomposition and the interaction of distribution movements and functional form structure, we now turn to a discussion of linearity of the aggregate function in μ. For this purpose, we refine Theorems 1 and 2 to:

Theorem 6: There exists an orthogonal decomposition of $\underline{p}$ which is minimal in θ_1;

$$\underline{p} = \frac{1}{N}\underline{i} + \sum_{i=1}^{N_1} e_i(\mu)\underline{\delta}_i + \sum_{k=1}^{N_2} f_k(\mu)\underline{\eta}_i + \sum_{j=1}^{N_3} g_j(\mu,\theta_1)\underline{\xi}_j$$

such that $N_1 + N_2 + N_3 \leq N-1$, each $e_i(\mu)$ is a linear function of μ, and if $\sum_{k=1}^{N_2} \alpha_k f_k(\mu) = c + d\mu$, then $\alpha_k = 0$ for all k. The subspaces $S_\eta = \text{SPAN}\left\{\underline{\eta}_k\right\}$ and $S_{\theta_1} = \text{SPAN}\left\{\xi_j\right\}$ are unique for all such decompositions. Moreover, if $\underline{x}$ represents a function $x = x(A)$, then $E(x)$ is a linear function of μ if

$$\underline{x} \in (S_\eta \oplus S_{\theta_1})^{\perp}$$

where $S_\eta \oplus S_{\theta_1} = \text{SPAN}\left\{\underline{\eta}_1,\ldots,\underline{\eta}_{N_2},\underline{\xi}_1,\ldots,\underline{\xi}_{N_3}\right\}$

Proof: See the Appendix.

Thus $E(x)$ is a linear function of μ iff $\underline{x}$ is orthogonal to all directions where either movement of the density is nonlinear in μ (via $f_k(\mu)$) or dependent upon θ_1 (via $g_j(\mu,\theta_1)$). Thus, $\underline{x}$ must be orthogonal to all directions of movement of $\underline{p}$ except those obeying LPM structure. Each of the constancy conditions referred to in the examples reduced a g_j function to one linear in μ, giving $E(x)$ in each case as a linear function of μ.

The final topic we address in this section is when there is more than one variable, say A and B, defined over the N possible consumer types. If $x = x(A,B)$, then when can $E(x)$ be written as a function only of $\mu_A = E(A)$ and $\mu_B = E(B)$? Here we assume $\theta = (\mu_A,\mu_B,\theta_1)$ so that $\theta_0 = (\mu_A,\mu_B)$ in the earlier notation.

The analysis of this case is the same as the previous case with one modification. Let $\underline{A}$ represent A and $\underline{B}$ represent B. As before, we define the direction $\underline{\tilde{A}}$ uniquely associated with μ_A. While a similar direction could be defined for μ_B, it would in general not be orthogonal to $\underline{\tilde{A}}$. To insure this, we define $\underline{\tilde{B}}$ as

$$\underline{\tilde{B}} = \underline{B} - \overline{B}\underline{i} - r\,\underline{\tilde{A}}$$

where $\overline{B} = \sum_{i=1}^{N} B_i/N$ and $r = \underline{B}'\underline{\tilde{A}}/\underline{\tilde{B}}'\underline{\tilde{A}}$.

The vectors $\underline{i}$, $\underline{\tilde{A}}$ and $\underline{\tilde{B}}$ are mutually orthogonal.

Now, in order to decompose the density vector $\underline{p}(\theta) = \underline{p}(\mu_A,\mu_B,\theta_1)$, we choose N-3 vectors $\underline{\delta}_3,\ldots,\underline{\delta}_N$ such that $\left\{\underline{i},\underline{\tilde{A}},\underline{\tilde{B}},\underline{\delta}_3,\ldots,\underline{\delta}_N\right\}$ is an orthogonal basis of R^N, and write

$$\underline{p}(\mu_A,\mu_B,\theta_1) = \frac{1}{N}\,\underline{i} + D_A(\mu_A)\underline{\tilde{A}} + D_B^*(\mu_A,\mu_B,\theta_1)\underline{\tilde{B}} + \sum_{i=3}^{N} d_i(\mu_A,\mu_B,\theta_1)\,\underline{\delta}_i$$

where $D_A(\mu_A) = (\mu_A-\overline{A})/\underline{\tilde{A}}'\underline{\tilde{A}}$ from the earlier development. $\underline{\tilde{A}}$ and $\underline{\tilde{B}}$ are unique directions associated with μ_A and μ_B, as

$$\mu_B = \underline{B}'\underline{p}(\mu_A,\mu_B,\theta_1) = \overline{B} + rD_A(\mu_A)\ \underline{\tilde{A}}'\underline{\tilde{A}} + D_B^*(\mu_A,\mu_B,\theta_1)\underline{\tilde{B}}'\underline{\tilde{B}}$$

$$= \overline{B} + r(\mu_A-\overline{A}) + D_B^*(\mu_A,\mu_B,\theta_1)\ \underline{\tilde{B}}'\underline{\tilde{B}}$$

so

$$D_B^*\ (\mu_A,\mu_B,\theta_1) = \frac{1}{\underline{\tilde{B}}'\underline{\tilde{B}}}\ \ (\mu_B - \overline{B} - r(\mu_A-\overline{A})) = D_B(\mu_A,\mu_B)$$

Having isolated the dependence of D_A and D_B on only μ_A and μ_B, we can proceed as before to consider a minimal decomposition in θ_1, producing an invariant subspace S_{θ_1}. If $\underline{x}$ represents the function $x = x(A,B)$, then $E(x) = \phi_0(\mu_A,\mu_B)$ iff $\underline{x}\,\varepsilon\, S_{\theta_1}^{\perp}$. We can also show the following result, which is analogous to Corollary 5:

Corollary 7:

(a) If $x = x(A,B)$ is a function such that $E(x) = \phi_0(\mu_A,\mu_B)$ for all densities obeying $\mu_A = E(A)$ and $\mu_B = E(B)$ then $x(A) = a + bA + cB$, a linear function of A and B, with $E(x) = \phi_0(\mu_A,\mu_B) = a + b\mu_A + c\mu_B$.

(b) If p is a density such that $E(x) = \phi_0(\mu_A,\mu_B)$ for all functions $x = x\ (A,B)$, then p depends only on μ_A and μ_B.

Proof: (a) and (b) are shown with precisely the same method of proof as in Corollary 5.

For (a), consider the density vector (N>3):

$$\underline{p}(\mu_A,\mu_B,d_3,\ldots,d_N) = \frac{1}{N}\,\underline{i} + D_A(\mu_A)\underline{\tilde{A}} + D_B(\mu_A,\mu_B)\ \underline{\tilde{B}} + \sum_{i=3}^{N} d_i\underline{\delta}_i$$

$E(x) = \phi_0(\mu_A,\mu_B)$, omitting $d_3,\ldots,d_N$ iff

$$\underline{x} = \overline{x}\underline{i} + b^*\underline{\tilde{A}} + c^*\underline{\tilde{B}}$$

$$= (\overline{x} + (rc^*-b)\overline{A} - c^*\overline{B})\underline{i} + (b^*-c^*r)\underline{A} + c^*\underline{B}$$

which corresponds to $x(A) = a + bA + cB$ where $a = \overline{x} + (rc^*-b)\overline{A} + c^*\overline{B}$, $b = (b^*-c^*r)$; $c = c^*$ are constants ($r = \underline{\tilde{B}}'\underline{\tilde{A}}/\underline{\tilde{A}}'\underline{A}$). (b) follows in the same way as Corollary 5(b).

Q.E.D.

Clearly these results generalize to the case where there are more than two predictor variables.

This completes our discussion of aggregate functions depending only on predictor variable means.

4.3 Aggregate Functions Depending only on Marginal Distributions

In this section we consider the problem of when an aggregate function depends only on the marginal distributions of the underlying predictor variables, where averaging occurs over the joint distribution of these variables. We analyze this problem for the case of discrete distributions, and our results are primarily all obtained by applying the theorems of the previous section. In order to make the correspondence between this section and the previous one, we require some additional notation.

In this problem A and B are assumed to be jointly distributed random variables. A has N possible (distinct) values A_i, $i = 1, \ldots, N$ and B has M possible values $B_j, j=1,\ldots,M$. The probability of (A_i,B_j) is denoted $p_{ij}(\theta)$, where θ is the vector of parameters determining the joint distribution over time. The marginal distribution of A is given by the probabilities:

$$\text{PROB } (A=A_i) = P_{Ai} = \sum_{j=1}^{M} p_{ij}(\theta), \; i=1,\ldots,N-1$$

and
$$\text{PROB } (A=A_N) = P_{AN} = 1 - \sum_{i=1}^{N-1} P_{Ai}$$

Similarly the marginal distribution of B is given by:

$$\text{PROB } (B=B_j) = P_{Bj} = \sum_{i=1}^{N} p_{ij}(\theta), \; j=1,\ldots,M-1$$

$$\text{PROB } (B=B_M) = P_{BM} = 1 - \sum_{j=1}^{M} P_{Bj}$$

We assume that $\theta = (P_{A1},\ldots,P_{AN-1},P_{B1},\ldots,P_{BM-1},\theta_1)$ where θ_1 is a set of other parameters determining the joint distribution. θ_1 is assumed nonempty; although we will later discuss the connection of our results to the case where A and B are independent, in which θ_1 is empty.

Here the behavioral function is $x = x(A,B)$ depending on both predictor variables. The problem we study here is the conditions under which the aggregate function $E(x) = \sum_{i,j} p_{ij}(\theta)\, x(A_i,B_j) = \phi_0(P_{A1},\ldots,P_{AN-1},P_{B1},\ldots,P_{BM-1})$, omitting θ_1 from $E(x)$.

In addition we denote by $\underline{i}_N$ the N vector of ones, by $\underline{i}_M$ the M vector of ones and by $\underline{i}$ the MN vector of ones. $\underline{e}_{jM}$ denotes the M vector with 1 in position j, zeros elsewhere and $\underline{e}_{iN}$ the N vector with 1 in position i, and zeros elsewhere. The MN vector of probabilities $\underline{p}(\theta)$ is formed as $\underline{p}(\theta) = (p_{1},(\theta),p_{1}(\theta),\ldots,p_{1M}(\theta),p_{21}(\theta),\ldots,p_{N1}(\theta),\ldots,p_{NM}(\theta))'$. Now, if we define the following MN vectors

$$\underline{p}_{Ai} = \underline{e}_{iN} \otimes \underline{i}_M \;\; ; \quad i=1,\ldots,N-1$$

$$\underline{p}_{Bj} = \underline{i}_N \otimes \underline{e}_{jM} \;\; ; \quad j=1,\ldots,M-1$$

we find that

$$P_{Ai} = \underline{\rho}_{Ai}'\underline{p}(\theta) \quad ; \quad i=1,\ldots,N-1$$

$$P_{Bj} = \underline{\rho}_{Bj}'\underline{p}(\theta) \quad ; \quad j=1,\ldots,M-1 \tag{4.9}$$

The vectors $\underline{\rho}_{Ai}$, $i=1,\ldots,N-1$ and $\underline{\rho}_{Bj}$, $j=1,\ldots,M-1$ will play the same role as the vectors $\underline{A}$ and $\underline{B}$ of the previous section. There is one major advantage to the structure of $\underline{\rho}_{Ai}$ and $\underline{\rho}_{Bi}$; seen by first orthogonally decomposing them as

$$\underline{\rho}_{Ai} = \underline{\rho}_{Ai}^{*} + \frac{1}{N}\underline{i} \quad ; \quad i=1,\ldots,N-1$$

$$\underline{\rho}_{Bj} = \underline{\rho}_{Bj}^{*} + \frac{1}{M}\underline{i} \quad ; \quad j=1,\ldots,M-1$$

and showing the following Proposition:

Proposition 8: If $s_1 \in \text{SPAN}\left\{\underline{\rho}_{Ai}^{*},\ldots,\underline{\rho}_{AN-1}^{*}\right\}$ and

$s_2 \in \text{SPAN}\left\{\underline{\rho}_{B1}^{*},\ldots,\underline{\rho}_{BM-1}^{*}\right\}$

then $s_1's_2 = 0$.

Proof: For each i, j, note that

$$\underline{\rho}_{Ai}^{*}{}'\underline{\rho}_{Bj}^{*} = (\underline{\rho}_{Ai}-\frac{1}{N}\underline{i})'(\underline{\rho}_{Bj}-\frac{1}{M}\underline{i})'$$

$$= \underline{\rho}_{Ai}'\underline{\rho}_{Bj} - \frac{1}{N}\underline{i}'\underline{\rho}_{Bj} - \frac{1}{M}\underline{\rho}_{Ai}'\underline{i} + \frac{1}{NM}\underline{i}'\underline{i}$$

$$= 1 - 1 - 1 + 1 = 0$$

Any linear combination of $\underline{\rho}_{Ai}^{*}$'s is therefore orthogonal to any linear combination of $\underline{\rho}_{Bj}^{*}$'s.

Q.E.D.

The implications of this lemma for orthogonally decomposing $\underset{\sim}{p}(\theta)$ are seen as follows: first define M + N-1 mutually orthogonal vectors $\underset{\sim}{i}$, $\tilde{\underset{\sim}{\rho}}_A$, ..., $\tilde{\underset{\sim}{\rho}}_{AN-1}$, $\tilde{\underset{\sim}{\rho}}_{B1}$, ..., $\tilde{\underset{\sim}{\rho}}_{BM-1}$ as

$$\tilde{\underset{\sim}{\rho}}_{A1} = \underset{\sim}{\rho}^*_{A1}$$

$$\tilde{\underset{\sim}{\rho}}_{Ai} = \underset{\sim}{\rho}^*_{Ai} - \sum_{k\leq i-1} r_{ki}\, \tilde{\underset{\sim}{\rho}}_{Ak} \qquad i=2,\ldots,N-1$$

and

$$\tilde{\underset{\sim}{\rho}}_{B1} = \underset{\sim}{\rho}^*_{B1}$$

$$\tilde{\underset{\sim}{\rho}}_{Bj} = \tilde{\underset{\sim}{\rho}}^*_{Bj} - \sum_{k\leq j-1} s_{kj}\, \tilde{\underset{\sim}{\rho}}_{Bj} \qquad i=2,\ldots,M-1$$

where $r_{ki} = (\tilde{\underset{\sim}{\rho}}_{Ak}{}'\underset{\sim}{\rho}_{Ai}/\tilde{\underset{\sim}{\rho}}_{Ak}{}'\tilde{\underset{\sim}{\rho}}_{Ak})$, $s_{kj} = (\tilde{\underset{\sim}{\rho}}_{Bk}{}'\underset{\sim}{\rho}_{Bj}/\tilde{\underset{\sim}{\rho}}_{Bk}\tilde{\underset{\sim}{\rho}}_{Bk})$, and from Lemma 9 there is no need to orthogonalize ρ_{Bj}, j=1, ...,M-1, with respect to $\tilde{\underset{\sim}{\rho}}_{Ai}$. To decompose the density, we choose MN - M - N+1 vectors $\underset{\sim}{\delta}_k$, k = M + N,...., MN, such that $\left\{\underset{\sim}{i}, \tilde{\underset{\sim}{\rho}}_{A1}, \ldots, \tilde{\underset{\sim}{\rho}}_{AN-1}, \tilde{\underset{\sim}{\rho}}_{B1}, \ldots, \tilde{\underset{\sim}{\rho}}_{BM-1}, \underset{\sim}{\delta}_{M+N}, \ldots, \underset{\sim}{\delta}_{MN}\right\}$ is an orthogonal basis of R^{MN}, and write $\underset{\sim}{p}(\theta)$ as

$$\underset{\sim}{p}(\theta) = \frac{1}{MN}\underset{\sim}{i} + \sum_{i=1}^{N-1} D^*_{Ai}(\theta)\underset{\sim}{\rho}^*_{Ai} + \sum_{j=1}^{M-1} D^*_{Bj}(\theta)\, \underset{\sim}{\rho}_{Bj} + \sum_{k=M+N}^{MN} d_k(\theta)\, \underset{\sim}{\delta}_k \qquad (4.10)$$

Solving out each of the equations in (4.9) yields:

$$D^*_{Ai}(\theta) = \frac{1}{\tilde{\rho}_{A1}{}'\tilde{\rho}_{A1}}\left(P_{A1} - \frac{1}{N}\right) = D_{A1}(P_{A1})$$

$$D^*_{Ai}(\theta) = \frac{1}{\tilde{\rho}_{Ai}{}'\tilde{\rho}_{Ai}}\left(P_{Ai} - \frac{1}{N} - \sum_{k\leq i-1} D_{Ak} r_{ik}\tilde{\underset{\sim}{\rho}}_{Ak}{}'\tilde{\underset{\sim}{\rho}}_{Ak}\right) \qquad (4.11)$$

$$= D_{Ai}(P_{Ai},\ldots,P_{Ai}) \qquad i=2,\ldots,N-1$$

$$D^*_{B1}(\theta) = \frac{1}{\tilde{\underset{\sim}{\rho}}_{B1}{}'\tilde{\underset{\sim}{\rho}}_{B1}}\left(P_{B1} - \frac{1}{N}\right) = D_{B1}(P_{B1})$$

$$D^*_{Bj}(\theta) = \frac{1}{\tilde{\underline{\rho}}_{Bj}'\tilde{\underline{\rho}}_{Bj}}\left(P_{Bj} - \sum_{k\leq j-1} D_{Bk} S_{ik}\, \tilde{\underline{\rho}}_{Bk}'\tilde{\underline{\rho}}_{Bk}\right)$$

$$= D_{Bj}(P_{B1},\ldots,P_{Bj}) \quad ; \quad j=1,\ldots,M-1$$

Thus, the coefficient D_{Ai} depends only on $P_{A1},\ldots,P_{Ai}$; the coefficient D_{Bj} depends only on $P_{B1},\ldots,P_{Bj}$. Thus $\tilde{\underline{\rho}}_{A1},\ldots,\tilde{\underline{\rho}}_{AN-1}$ are directions corresponding to the parameters $P_{A1},\ldots,P_{AN-1}$ and $\tilde{\underline{\rho}}_{B1},\ldots,\tilde{\underline{\rho}}_{M-1}$, are directions corresponding to the parameters $P_{B1},\ldots,P_{BM-1}$. This is true for all joint distributions with $P_{A1},\ldots,P_{AN-1},P_{AN},P_{B1},\ldots,P_{BM-1},P_{BM}$ as marginal probabilities. Moreover, as can be checked from (4.11), each D_{Ai} and D_{Bj} coefficient is linear in its respective arguments.

With this background, we can clearly now proceed as in the previous section, by writing (4.10) in minimal (in θ_1) form, defining S_{θ_1}, and characterizing the precise structure for a behavioral function $x = x(A,B)$ to allow $E(x) = \phi_0(P_{A1},\ldots,P_{AN-1},P_{B1},\ldots,P_{BM-1})$. While this analysis is quite valid, it does not exhibit any particular special properties of the appearance of marginal distributions versus that of several means as the arguments of an aggregate function. The following result exhibits a different form than the previous theorems, but it is actually just a translation of Corollary 7 into the current framework.

Corollary 9: Given Assumption 1, $x = x(A,B)$ is a function such that $E(x) = \phi_0(P_{A1},\ldots,P_{AN-1},P_{B1},\ldots,P_{BM-1})$ for all joint distributions of A and B with $P_{A1},\ldots,P_{AN}$ as the marginal distribution of A and $P_{B1},\ldots,P_{BM}$ that of B, then $x(A,B) = x_A(A) + x_B(B)$, i.e. $x(A,B)$ must be additive in A and B. If $x(A,B)$ is such that

$E(x) = \phi_0(P_{A1},\ldots,P_{AN-1})$ for all such distributions, then $x(A,B) = x_A(A)$, i.e. $x(A,B)$ depends only on A. Similarly, if $E(x) = \phi(P_{B1},\ldots,P_{BM-1})$ for all such distributions, then $x(A,B) = x_B(B)$

Proof: Consider the density vector in the form (4.10) given as

$$\underline{\rho}(P_{A1},\ldots,P_{AN-1},P_{B1},\ldots,P_{BM-1},d_{M+N},\ldots,d_{MN})$$

$$= \frac{1}{MN}\underline{i} + \sum_{i=1}^{N-1} D_{Ai}(P_{A1},\ldots,P_{Ai})\underline{\tilde{\rho}}_{Ai} + \sum_{j=1}^{M-1} D_{Bj}(P_{B1},\ldots,P_{Bj})\underline{\tilde{\rho}}_{Bj}$$

$$+ \sum_{k=M+N}^{MN} d_k \underline{\delta}_k$$

If $\underline{x}$ is the vector representing $x = x(A,B)$, then $E(x) = \phi_0(P_{A1},\ldots,P_{AN-1},P_{B1},\ldots,P_{BM-1})$ if and only if:

$$\underline{x} = \bar{x}\underline{i} + \sum_{i=1}^{N-1} u_i \underline{\tilde{\rho}}_{A_i} + \sum_{j=1}^{M-1} v_j \underline{\tilde{\rho}}_{Bj}$$

$$= a\underline{i} + \sum_{i=1}^{N-1} b_i \underline{\rho}_{Ai} + \sum_{j=1}^{M-1} c_j \underline{\rho}_{Bj}$$

where b_i, $i=1,\ldots N-1$ are constants arrived at by solving for each $\underline{\tilde{\rho}}_{Ai}$ in terms of $\underline{\rho}_{Ai}$, $i=1,\ldots,N-1$, and c_j, $j=1,\ldots, M-1$ are arrived at by solving each $\underline{\tilde{\rho}}_{Bj}$ in terms of $\underline{\rho}_{Bj}$, $j=1,\ldots M-1$. If we define $x_A(A_i) = b_i$, $i=1,\ldots,N-1$. $x_A(A_N) = 0$, $x_B(B_j) = c_j$, $j=1,\ldots,M-1$, $x_B(B_M) = 0$, then we have that

$$x(A,B) = a + x_A(A) + x_B(B)$$

and since the constant a may be absorbed in the definition of $x_A(A)$ or $x_B(B)$, this can be written as:

$$x(A,B) = x_A(A) + x_B(B)$$

The other statements follow this line of proof exactly.

Q.E.D.

Proposition 8 serves only as a computational advantage for this problem, and does not play a role in the proof of Corollary 9. The appearance of orthogonality between the natural directions for $P_{Ai}, i=1,\ldots,N-1$ and those of $P_{Bj}, j=1,\ldots,M-1$, coincides with the property that any function only of A will have an expectation dependent only on $P_{Ai}, i=1,\ldots,N-1$, with the analogous property for functions dependent only on B.

We now illustrate the decomposition structure for joint distributions with two examples.

Example 7 (N=2,M=2):

Let A and B have possible values $A_1 = 1$, $A_2 = 0$, $B_1 = 1$, $B_2 = 0$. If the joint density vector is written $\underline{p}(\theta) = (p_{11}, p_{12}, p_{21}, p_{22})'$, then we have $P_{A1} = p_{11} + p_{12}$, $P_{B1} = p_{11} + p_{21}$. We reparameterize $\underline{p}(\theta)$ via P_{A1}, P_{B1} and $R = \mathrm{Cov}(A,B)$ as $\underline{p}(P_{A1}, P_{B1}, R) = (P_{A1}P_{B1} + R,\ P_{A1}(1-P_{B1}) - R,\ P_{B1}(1-P_{A1}) - R,\ (1-P_{A1})(1-P_{B1}) + R)'$. Here $\underline{\rho}_{A1} = (1,1,0,0)'$; $\underline{\rho}_{B1} = (1,0,1,0)'$; so $\underline{\tilde{\rho}}_{A1} = 1/2(1,1,-1,-1)'$; $\underline{\tilde{\rho}}_{B1} = 1/2(1,-1,1,-1)'$. Now, if $\underline{\delta} = 1/4(1,-1,-1,1)'$ then $\{\underline{i}, \underline{\tilde{\rho}}_{A1}, \underline{\tilde{\rho}}_{B1}, \underline{\delta}\}$ is an orthogonal basis of R^4, and we write:

$$\underline{p}(P_{A1}, P_{B1}, R) = \underline{i} + (P_{A1} - 1/2)\underline{\tilde{\rho}}_{A1} + (P_{B1} - 1/2)\underline{\tilde{\rho}}_{B1}$$

$$+ (4P_{A1}P_{B1} + 4R - 2P_A - 2P_B + 1)\,\underline{\delta}$$

Now, if $x = x(A,B)$ is a function of A and B represented by the vector $\underline{x}$, then $E(x)$ is not dependent on R if $\underline{x}'\underline{\delta} = 0$. In this case $\underline{x} = \bar{x}\underline{i} + u_1\underline{\tilde{p}}_{A1} + v_1\underline{\tilde{p}}_{B1}$ which corresponds to the behavioral function $x(A,B) = a + bA + cB$, where $a = \bar{x} - \frac{u_1+v_1}{2}$, $b = u_1$, $c = v_1$. Note that if R is constant, $E(x) = \phi_0(P_{A1},P_{B1})$ for all functions $x = x(A,B)$ and unless $x(A,B)$ is additive in A and B, $E(x)$ will contain a nonlinear term $(P_{A1}P_{B1})$ in P_{A1} and P_{B1}.

Example 8 (N=3,M=3):

Here A has possible values A_1, A_2, A_3 and B has possible values B_1, B_2 and B_3. If $\underline{p}(\theta) = (p_{11},p_{12},p_{13},p_{21},p_{22},p_{23},p_{31},p_{32},p_{33})'$, then $P_{A1} = p_{11} + p_{12} + p_{13}$, $P_{A2} = p_{21}, + p_{22} + p_{23}$, $P_{B1} = p_{11} + p_{21} + p_{31}$ and $P_{B2} = p_{12} + p_{22} + p_{32}$. We can reparameterize $\underline{p}(\theta)$ with P_{A1}, P_{A2}, P_{B1}, P_{B2} and four additional parameters R_{11}, R_{12}, R_{21}, R_{22} as:

$$\underline{p}(P_{A1},P_{A2},P_{B1},P_{B2},R_{11},R_{12},R_{21},R_{22}) = \begin{bmatrix} P_{A1}P_{B1} + R_{11} \\ P_{A1}P_{B2} + R_{12} \\ P_{A1}(1-P_{B1}-P_{B2}) - R_{11} - R_{12} \\ P_{A2}P_{B1} + R_{21} \\ P_{A2}P_{B2} + R_{22} \\ P_{A2}(1-P_{B1}-P_{B2}) - R_{21} - R_{22} \\ (1-P_{A1}-P_{A2})P_{B1} - R_{11} - R_{21} \\ (1-P_{A1}-P_{A2})P_{B2} - R_{12} - R_{22} \\ (1-P_{A1}-P_{A2})(1-P_{B1}-P_{B2}) + R_{11} + R_{12} + R_{21} + R_{22} \end{bmatrix}$$

The reader can easily verify that:

$$\underset{\sim}{\rho}_{A1} = (1,1,1,0,0,0,0,0,0) \quad ; \quad \underset{\sim}{\rho}_{A2} = (0,0,0,1,1,1,0,0,0)'$$

$$\underset{\sim}{\rho}_{B1} = (1,0,0,1,0,0,1,0,0)' \quad ; \quad \underset{\sim}{\rho}_{B2} = (0,1,0,0,1,0,0,1,0)'$$

from which are derived:

$$\underset{\sim}{\tilde{\rho}}_{A1} = 1/3(2,2,2,-1,-1,-1,-1,-1,-1)' \quad ; \quad \underset{\sim}{\tilde{\rho}}_{A2} = 1/2(0,0,0,1,1,1,-1,-1,-1)'$$

$$\underset{\sim}{\tilde{\rho}}_{B1} = 1/3(2,-1,-1,2,-1,-1,2,-1,-1)' \quad ; \quad \underset{\sim}{\tilde{\rho}}_{B2} = 1/2(0,1,-1,0,1,-1,0,1,-1)'$$

The coefficients D_{A1}, D_{A2}, D_{B1}, D_{B2} of $\underset{\sim}{p}(\theta)$ are given as:

$$D_{A1} = \frac{P_{A1}}{2} - \frac{1}{6} \quad ; \quad D_{A2} = \frac{2P_{A2} + P_{A1} - 1}{3}$$

$$D_{B1} = \frac{P_{B1}}{2} - \frac{1}{6} \quad ; \quad D_{B2} = \frac{2P_{B2} + P_{B1} - 1}{3}$$

Because of the size and complexity of the vectors in this example, we do not explicitly present the decomposition or illustrate restricted density movements from it. As according to Corollary 9, if $x = x(A,B)$ is a behavioral function, then $E(x) = \phi_0(P_{A1},P_{A2},P_{B1},P_{B2})$ if and only if x is additive in A and B, when $R_{11},R_{12},R_{21},R_{22}$ are changing parameters. However, if the distribution movement is restricted so that for some i,j,R_{ij} is constant, then $x(A,B)$ may be non-additive for particular values of A and B, while $E(x)$ is independent of the remaining $R_{i'j'}$ parameters. The impact of these restrictions is summarized in the following table:

Parametric Aspect Constant	Nonadditive Function Value
R_{11}	$x(A_3,B_3)$
R_{12}	$x(A_3,B_2)$
R_{21}	$x(A_2,B_3)$
R_{22}	$x(A_2,A_2)$

This table is interpreted as in: if $E(x) = \phi_0(P_{A1},P_{A2},P_{B1},P_{B2})$ and R_{21} is constant, then $x(A,B)$ must be of the form

$$x = x_A(A) + x_B(B) + x_{AB}(A,B) \quad ,$$

where $x_{AB}(A,B) = c$ if $A = A_2$, $B = B_3$, 0 otherwise.

In each of these examples, nonadditive $x(A,B)$ functions allowed $E(x)$ to omit all parameters other than the marginal probabilities only when certain of the remaining parameters were constant. For parameter omission issues, independence of A and B is only a means of guaranteeing that the remaining distributional parameters are constant.[29] Thus, in Example 7, if R is a non-zero constant, $E(x) = \phi_0(P_{A1},P_{B1})$. The independent case (R=0) provides no other structure relevant to omitting parameters. In Example 8, independence of A and B translates to $R_{11} = R_{12} = R_{21} = R_{22} = 0$, giving $E(x) = \phi_0(P_{A1},P_{A2},P_{B1},P_{B2})$ for all functions $x = x(A,B)$.

We close this section by considering the structure under which $E(x)$ is a function only of μ_A and μ_B where $\mu_A = E(A)$, $\mu_B = E(B)$. If we define $\underline{A}$ and $\underline{B}$ (vectors representing the functions $(A,B)\to A, (A,B)\to B$) then we have that $\underline{A}-\bar{A}\underline{i} = \sum_{i=1}^{N-1} A_i \underline{\rho}^*_{Ai}$, $\underline{B}-\bar{B}\underline{i} = \sum_{j=1}^{N-1} B_j \underline{\rho}^*_{Bj}$ and $(\underline{A}-\bar{A}\underline{i})'(\underline{B}-\bar{B}\underline{i}) = 0$. Repeating the construction in the proof of Corollary 9, we see that $E(x)$ will depend only on μ_A and μ_B if and only if $b_i = bA_i$, $c_j = cB_j$, giving $x(A,B)$ as $x(A,B) = a + bA_i + cB_j$. Clearly this restriction is in accordance with Corollary 7.

4.4 Omitting Segmented Distribution Detail

In the two previous sections, we have found that parameters could be omitted from E(x) for general distribution movements only if the function x had a linear structure. Nonlinear x structures allow parameter omission only when particular restrictions were placed on the movement of the distribution.

In this section we consider a particular type of restricted distribution movement, namely when the underlying population is composed of two segments. As in the Introduction, these segments could correspond to urban individuals and rural individuals, with the urban segment parameterized by mean income for urban individuals, and the rural segment parameterized by rural mean income. Our interest here is in when this segmented detail is irrelevant to an aggregate function, i.e. when E(x) depends only on overall mean income. We first analyze this case for general problems obeying Assumptions 1 and 2, and then illustrate by presenting the decomposition for discrete densities applied to this problem.

Formally, suppose that the density of the population has the form:

$$p(A|\lambda,\mu_1,\mu_2) = \lambda p_1(A|\mu_1) + (1-\lambda)p_2(A|\mu_2) \tag{4.12}$$

where each segment density p_i is parameterized by μ_i, i=1,2 and λ is the proportion of individuals in segment 1. The overall parameter of interest here is $\mu = \lambda\mu_1 + (1-\lambda)\mu_2$, and so we reparameterize the density (4.12) to:

$$p(A|\mu,\lambda,\mu_1) = \lambda p_1(A|\mu_1) + (1-\lambda)p_2(A|\frac{\mu-\lambda\mu_1}{1-\lambda}) \tag{4.13}$$

If x = x(A) is the individual behavioral function, then the aggregate function is:

$$E(x) = \phi(\mu,\lambda,\mu_1) = \lambda E_1(x|\mu_1) + (1-\lambda)E_2(x|\frac{\mu-\lambda\mu_1}{1-\lambda}) \tag{4.14}$$

where $E_1(x|\mu_1) = \int x(A)p_1(A|\mu_1)dA$, $E_2(x|\mu_2) = \int x(A)p_2(A|\mu_2)dA$

Our interest is in the conditions under which $E(x) = \phi_0(\mu)$, omitting the parameters λ, μ_1. In particular, are there any nonlinear behavioral functions which allow $E(x) = \phi_0(\mu)$ when distribution movement is in segmented form?

The answer to this question is largely negative, as given by the following theorem:

Theorem 10: Assume Assumption 1, the density is of the form (4.14) and Assumption 2 holds when applied to the behavioral function $x = x(A)$ and each of the segment densities p_1 and p_2. Then $E(x) = \phi_0'(\mu,\lambda)$ if and only if

$$E_1(x|\mu_1) = c_1 + b\mu_1$$

and (4.15)

$$E_2(x|\mu_2) = c_2 + b\mu_2$$

where c_1, c_2 and b are constants. $E(x) = \phi_0(\mu)$ if and only if $c_1 = c_2$.

Proof: We apply eqn. (4.1A) with $\theta = (\mu,\lambda,\mu_1)$ and $\theta_1 = \mu_1$, giving:

$$0 = \frac{\partial E(x)}{\partial \mu_1} = \lambda \int x(A) \left[\frac{\partial p_1}{\partial \mu_1} - \frac{\partial p_2}{\partial \mu_2} \right] dA$$

which is rewritten as:

$$\frac{\partial E_1(x|\mu_1)}{\partial \mu_1} = \int x(A) \frac{\partial p_1}{\partial \mu_1} dA = \int x(A) \frac{\partial p_2}{\partial \mu_2} dA = \frac{\partial E_2(x|\mu_2)}{\partial \mu_2}$$

Since the LHS depends on μ_1, and the RHS depends on $\mu_2 = \left[\mu - \lambda\mu_1 / 1-\lambda \right]$, then each of the above derivatives $\frac{\partial E_1(x)}{\partial \mu_1}$ $\frac{\partial E_2(x)}{\partial \mu_2}$ must be constant, equal to b, say.

Thus, we must have (4.15), i.e.

$$E_1(x|\mu_1) = c_1 + b\mu_1$$

$$E_2(x|\mu_2) = c_2 + b\mu_2$$

With this structure

$$E(x) = \lambda c_1 + (1-\lambda)c_2 + b\mu$$

Obviously, $E(x) = \phi_0(\mu)$, omitting λ, if and only if $c_1 = c_2$.

Q.E.D.

Thus, in order to ignore segmented distribution detail, we must have the same linear aggregate function over each distribution segment. This requires the conditions of Section 4.2 on each segment. Note that we have not assumed that $\mu_1 = E_1(A|\mu_1)$, $\mu_2 = E_2(A|\mu_2)$ (with $\mu = E(A)$), although in empirical applications this structure will usually exist.

For illustration, we now consider p to be a discrete distribution, and show the decomposition relevant to this problem. This presentation will show the role of conditional distributions in the orthogonal decomposition. We begin with the notation and set-up of this problem. Here we assume that $\mu = E(A)$, $\mu_1 = E_1(A|\mu_1)$ and $\mu_2 = E_2(A|\mu_2)$.

Assume that there are M+N types of consumers, with M in segment 1 and N in segment 2. We begin by representing the conditional distributions $p_1(A|\mu_1)$ and $p_2(A|\mu_2)$ as M+N vectors by $\underline{p}_1(\mu_1) = (p_{11}(\mu_1),\ldots,p_{1M}(\mu_1),0,0,\ldots,0)'$ and $\underline{p}_2(\mu_2) = (0,\ldots,0,p_{21}(\mu_2),\ldots,p_{2N}(\mu_2))'$, where only the first M components of $\underline{p}_1(\mu_1)$ are nonzero and only the last N components of $\underline{p}_2(\mu_2)$ are nonzero. The M+N vector of probabilities for the entire population is then [30]

$$\underline{p}(\mu_1,\mu_2,\lambda) = \lambda\underline{p}_1(\mu_1) + (1-\lambda)\,\underline{p}_2(\mu_2) \tag{4.16}$$

Define $\underline{i}$ as the M+N vector of ones, $\underline{i}_{M0}$ as the M+N vector with ones in the first M positions and zeros elsewhere, and $\underline{i}_{0N} = \underline{i} - \underline{i}_{M0}$. Let $\underline{A}$ be the vector representing the possible values of A, and define $\underline{A}_1, \underline{A}_2$ as the vectors $\underline{A}_1 = (A_1, \ldots, A_M, 0, \ldots, 0)$ and $A_2 = (0, \ldots, 0, A_{M \times 1}, \ldots, A_{M+N})'$ so that

$$\underline{A} = \underline{A}_1 + \underline{A}_2 \tag{4.17}$$

Our first step is to decompose the conditional densities, paying special attention to the normalizing vectors $\underline{i}_{M0}$ and $\underline{i}_{0N}$ and the directions $\underline{A}_1$ and $\underline{A}_2$. As in previous sections, we decompose $\underline{A}_1$ and $\underline{A}_2$ orthogonally as:

$$\underline{A}_1 = \bar{A}_1\ \underline{i}_{M0} + \tilde{\underline{A}}_1$$

$$\underline{A}_2 = \bar{A}_2\ \underline{i}_{0N} + \tilde{\underline{A}}_2$$

where $\bar{A}_1 = \sum_{i=1}^{M} A_i/M$, $\bar{A}_2 = \sum_{j=1}^{N} A_{M+j}/N$. We decompose $\underline{p}_1(\mu_1)$ by choosing M-2 vectors $\underline{\delta}_{i0}$, i=3, ..., M, such that $\left\{ i_{M0}, \tilde{\underline{A}}_1, \underline{\delta}_{30}, \ldots; \delta_{M0} \right\}$ is an orthogonal set of vectors, and each $\underline{\delta}_{i0}$ has zeros for the last N components. $\underline{p}_1(\mu_1)$ appears as:

$$\underline{p}_1(\mu_1) = \frac{1}{M}\underline{i}_{M0} + D_1(\mu_1)\tilde{\underline{A}}_1 + \sum_{i=3}^{M} d_i(\mu_1)\underline{\delta}_{i0}$$

where $D_1(\mu_1) = (\mu_1 - \bar{A}_1)/\tilde{\underline{A}}_1'\tilde{\underline{A}}_1$. Similarly, $\underline{p}_2(\mu_2)$ can be written as:

$$\underline{p}_2(\mu_2) = \frac{1}{N}\underline{i}_{0N} + D_2(\mu_2)\ \tilde{\underline{A}}_2 + \sum_{j=3}^{N} f_j(\mu_2)\underline{\delta}_{0j}$$

with $\left\{ \underline{i}_{0N}, \tilde{\underline{A}}_2, \underline{\delta}_{03}, \ldots, \underline{\delta}_{0N} \right\}$ an orthogonal set of vectors such that the first M components of each $\underline{\delta}_{0j}$ are zero, and with $D_2(\mu_2) = (\mu_2 - \bar{A}_2)/\tilde{A}_2'\tilde{A}_2$. The overall density vector is then (from (4.16)):

$$\underline{p}(\mu_1,\mu_2,\lambda) = \frac{\lambda}{M}\underline{i}_{OM} + \frac{(1-\lambda)}{N}\underline{i}_{ON} + \lambda D_1(\mu_1)\tilde{\underline{A}}_1$$
$$+ (1-\lambda)D_2(\mu_2)\tilde{\underline{A}}_2 + \sum_{i=3}^{M} \lambda d_i(\mu_1)\underline{\delta}_{io} + \sum_{j=3}^{N} (1-\lambda)f_j(\mu_2)\underline{\delta}_{oj} \quad (4.18)$$

Recall that our interest in this problem is in aggregate functions which depend only on $E(A) = \mu = \lambda\mu_1 + (1-\lambda)\mu_2$ or only on μ and λ. Noting that

$$\mu = \underline{A}'\underline{p}(\mu_1,\mu_2,\lambda)$$

$$\lambda = \underline{i}_{MO}'\,\underline{p}(\mu_1,\mu_2,\lambda)$$

we can make (4.18) useful by transforming its basis to one representing the $\underline{A}$ and $\underline{i}_{MO}$ directions, as well as the initial normalizing vector $\underline{i}$. As before, we define the orthogonal vectors:

$$\tilde{\underline{A}} = \underline{A} - \bar{A}i \quad ; \quad \bar{A} = \sum_{i=1}^{M+N} A_i/M+N$$

$$\tilde{\underline{i}}_{MO} = \underline{i}_{MO} - \frac{M}{M+N}\underline{i} - S\tilde{\underline{A}} \quad ; \quad S = \frac{MN}{M+N}\,\frac{\bar{A}_1-\bar{A}_2}{\tilde{\underline{A}}'\tilde{\underline{A}}}$$

As we wish to transform the coefficients of (4.18) to those of $\underline{i}$, $\underline{\tilde{A}}$ and $\underline{\tilde{i}}_{MO}$, we first note that $\underline{i}$, $\underline{\tilde{A}}$, $\underline{\tilde{i}}_{MO} \in$ SPAN $[\underline{i}_{MO}, \underline{i}_{ON}, \tilde{A}_1, \tilde{A}_2]$ and if we define

$$\underline{\xi} = (1 - r)\underline{\tilde{A}}_1 - r\underline{\tilde{A}}_2 ; \; r = \frac{\underline{\tilde{A}}_1'\underline{\tilde{A}}_1}{\underline{\tilde{A}}_1'\underline{\tilde{A}}_1 + \underline{\tilde{A}}_2'\underline{\tilde{A}}_2}$$

then $[\underline{i}, \underline{\tilde{A}}, \underline{\tilde{i}}_{MO}, \underline{\xi}]$ is an orthogonal set of vectors, with SPAN$[\underline{i}, \underline{\tilde{A}}, \underline{\tilde{i}}_{MO}, \underline{\xi}]$ = SPAN $[\underline{i}_{MO}, \underline{i}_{ON}, \underline{\tilde{A}}_1, \underline{\tilde{A}}_2]$. Transforming the coefficients of (4.18) to correspond to these new vectors gives

(4.19)
$$\underline{p}(\mu_1, \mu_2, \lambda) = \frac{1}{M + N}\underline{i} + D(\mu)\underline{\tilde{A}} + D_i(\mu, \lambda)\underline{\tilde{i}}_{MO}$$

$$+ D_\xi(\mu_1, \mu_2, \lambda)\underline{\xi} + \sum_{i=3}^{M} \lambda d_i(\mu_1)\underline{\delta}_{io} + \sum_{j=3}^{N} (1 - \lambda) f_j(\mu_2)\underline{\delta}_{oj}$$

where
$$D(\mu) = \frac{\lambda\mu_1 + (1 - \lambda)\mu_2 - \bar{A}}{\underline{\tilde{A}}'\underline{\tilde{A}}} = \frac{\mu - \bar{A}}{\underline{\tilde{A}}'\underline{\tilde{A}}}$$

$$D_i(\mu, \lambda) = \frac{\lambda - \frac{M + N}{M} - S(\mu - \bar{A})}{\underline{\tilde{i}}_{MO}'\underline{\tilde{i}}_{MO}}$$

$$D_\xi(\mu_1, \mu_2, \lambda) = \left[\frac{\lambda(\mu_1 - \bar{A}_1)}{\underline{\tilde{A}}_1'\underline{\tilde{A}}_1} - \frac{(1 - \lambda)(\mu_2 - \bar{A}_2)}{\tilde{A}_2'\tilde{A}_2}\right]$$

(4.19) is in the proper form for considering parameter omission in this problem.

Suppose that x = x(A) is a behavioral function represented by a vector $\underline{x}$. The vectors $\underline{i}$, $\underline{\tilde{A}}$, $\underline{\tilde{i}}_{MO}$ and $\underline{\xi}$ above pertain to special functional forms as follows. If $\underline{x}$ is collinear with $\underline{i}$, then x = x(A) is constant. If $\underline{x} \in$ SPAN $[i, \tilde{A}]$ = SPAN $[\underline{i}, \underline{A}]$, then x(A) is a linear function of A common to both segments. If $\underline{x} \in$ SPAN $[\underline{i}, \underline{\tilde{A}}, \underline{\tilde{i}}_{MO}]$ = SPAN $[\underline{i}, \underline{A}, \underline{i}_{MO}]$, then x(A) is a linear function of A, with a different constant term in each segment. Finally, if $\underline{x} \in$ SPAN $[\underline{i}, \underline{\tilde{A}}, \underline{i}_{MO}, \underline{\xi}]$ = SPAN $[\underline{i}_{MO}, \underline{i}_{ON}, \underline{\tilde{A}}_1, \underline{\tilde{A}}_2]$ then x(A) is a linear function of A for each segment, with differing constants and slope coefficients. Thus, adding each one of the above vectors to the composition of $\underline{x}$ adds a natural degree of complexity to the functional structure of $x = x(A)$.

In addition, if the forms of $d_i(\mu_1)$ and $f_j(\mu_2)$ in (4.18) and (4.19) are unrestricted, then from previous developments $E(x) = \phi_0(\mu)$ iff $\underline{x} \in$ SPAN $\underline{i}, \underline{\tilde{A}}]$, i.e. x is a linear of A. Similarly, in this case $E(x) = \phi_0'(\mu, \lambda)$ iff $\underline{x} \in$ SPAN $[\underline{i}, \underline{\tilde{A}}, \underline{\tilde{i}}_{MO}]$. Any difference in slope coefficients of x(A) across segments will violate $E(x) = \phi_0'(\mu, \lambda)$, since D_ξ cannot be written solely as a function of μ and λ.

$x = x(A)$ can depart from the above linearity assumptions in two ways so as to guarantee $E(x) = \phi_0'(\mu, \lambda)$. First, $\underline{x}$ could lie in the subspace orthogonal to SPAN $[\underline{\delta}_{30}, \ldots, \underline{\delta}_{MO}, \underline{\delta}_{03}, \ldots, \underline{\delta}_{ON}]$, in which case either x(A) is linear or has support in directions with zero coefficients in $\underline{p}(\mu_1, \mu_2, \lambda)$. Otherwise the distribution must be restricted, i.e. if

$$\bar{x} = a\underline{i} + b\underline{\tilde{A}} + c\underline{\tilde{i}}_{MO} + \sum_{i=3}^{M} u_i \delta_{io} + \sum_{j=3}^{M} v_j \delta_{jo}$$

and $E(x) = \phi_0'(\mu, \lambda)$, then

$$\lambda \sum_{i=3}^{M} u_i d_i(\mu_1) + (1 - \lambda) \sum_{j=3}^{N} v_j f_j(\mu_2) = F(\mu, \lambda)$$

which implies by Jensen's equality that [31]

$$\sum_{i=3}^{M} u_i d_i(\mu_1) = h_1 + h\mu_1$$

$$\sum_{j=3}^{N} v_j f_j(\mu_2) = h_2 + h\mu_2$$

so that $d_i(\mu_1)$ and $f_j(\mu_2)$ must be restricted. In particular, with reference to Theorem 6, if we define a minimal nonlinear coefficient subspace S_n^1 and S_n^2 for each of $\underline{p}_1$ and $\underline{p}_2$, then if $d_i(\mu)$ is a coefficient of $\underline{\delta}_{io} \varepsilon S_n^1$, then we must have the corresponding $\mu_i = 0$. Similarly we must have $v_j = 0$ if $\delta_{oj} \varepsilon S_n^2$. In addition, in general if $x = x(A)$ is a linear function of A on one segment, then $x = x(A)$ cannot be a nonlinear function on the other segment if $E(x) = \phi_0'(\mu, \lambda)$. These results provide a strong underpinning of Theorem 10 for discrete distributions.

5. CHANGING DOMAIN OF PREDICTOR VARIABLES

All of the previous analysis has assumed that the domain of the underlying variable A is constant over time. Thus, attention was centered on the effects of compositional changes in distribution on the aggregate variable E(x). In this section we briefly discuss changes in both the composition of the distribution and the domain over which it operates, to show additional complexities arising from varying domain.

In the case of discrete distributions, the impact of varying domain occurs through the vector representations $\underline{A}$ and $\underline{x}$. Previously, these vectors were assumed constant, but under varying domain of A, $\underline{x}$ and $\underline{A}$ would change through time, possibly even in their order (i.e. the number of possible values of A). Consequently, the above analysis, which concentrates solely on the coefficients of the orthogonally decomposed density vector $\underline{p}(\theta)$, is not strictly applicable to this problem.

For practical problems, the ways in which the domain of the predictor variable can change can be quite complex, and thus studying their effect on the aggregate variable E(x) may require different techniques for different specific applications. Below we present a model which allows study of certain domain changing situations, but by no means all possible situations. However, this model may suffice for a large range of easily parameterized movements in the predictor variable domain, and thus may serve to aid analysis of many specific applications. As an example, we apply this model to perhaps the simplest mode of changing domain; namely translating and scaling.

We begin by assuming that there is a reference variable $a \epsilon \Omega$, where Ω is fixed domain over all time periods. a is distributed with density $p(a|\theta)$, where θ is a vector of parameters here indicating compositional change in the distribution of a. The actual predictor variable A is related to a via a function

$A = f(a,\chi)$, where χ is a vector of parameters which change over time. The domain of A, denoted $\Sigma_\chi = \{f(a,\chi)\,|\,a\varepsilon\Omega\}$, varies over time with χ. f is assumed to be invertible in a for any fixed value of χ, and differentiable in both a and the components of χ.

As before, our primary interest is in the expectation of $x = x(A)$, where x is now assumed differentiable in A. The aggregate function $E(x)$ now depends on both θ and χ and appears as

$$\begin{aligned} E(x) &= \int_\Omega x(f(a,\chi))p(a|\theta)da \\ &= \phi(\theta,\chi) \end{aligned} \tag{5.1}$$

Now, if θ_i is a component of θ, $E(x)$ does not depend on θ_i if $\frac{\partial E(x)}{\partial\theta_i} = 0$ for all χ and θ. Assuming that derivatives can be passed under the integral (5.1), this condition appears as

$$\begin{aligned} 0 = \frac{\partial E(x)}{\partial\theta_i} &= \int_\Omega x(f(a,\chi))\frac{\partial p}{\partial\theta_i}\, da \\ &= \mathrm{Cov}_a\left(x(f(a,\chi)),\ \frac{\partial \ln p}{\partial\theta_i}\right) \end{aligned} \tag{5.2}$$

A is distributed over Σ_χ with density $p_A(A|\theta, \chi) = p(f^{-1}(A, \chi)|\theta)\left|\frac{\partial f^{-1}}{\partial A}\right|$,

and so (5.2) can be rewritten as

$$0 = \frac{\partial E(x)}{\partial \theta_i} = \mathrm{Cov}_A(x(A)\ \frac{\partial \ln p_A}{\partial \theta_i}) \qquad (5.3)$$

This condition is exactly analogous to (4.1)b), and so omitting a compositional change parameter θ_i requires the same condition as in Section 4.1, and as before can be studied empirically by examining the covariance (5.3) (or (4.1)b).

In order to discuss omission of the parameters determining the domain of A, suppose that χ_j is a component of χ. E(x) does not depend on χ_j if for all θ,χ

$$0 = \frac{\partial E(x)}{\partial \chi_j} = \int_\Omega \frac{\partial x}{\partial A}\frac{\partial f}{\partial \chi_j}\, p(a|\theta)da = E_a(\frac{\partial x}{\partial A}\frac{\partial f}{\partial \chi_j}) \qquad (5.4)$$

This is the general condition for omitting a domain parameter χ_j. If $E_a(\frac{\partial f}{\partial \chi_j}) = 0$, as when E(A) does not depend on χ_j, then (5.4) can be written in covariance form as

$$0 = \mathrm{Cov}(\frac{\partial x}{\partial A}\frac{\partial f}{\partial \chi_j}) \qquad (5.5)$$

Conditions (5.4) and (5.5) can be used to find conditions for omitting a domain prarameter χ_j whenever there exists a reference random variable a and an invertible function f. For illustration we consider the case where the domain of A changes by translating or scaling, with a fixed reference density.

Example 9 (Translation and Scaling):

Suppose that a is a real random variable, $\Omega = (c, d)$ is an interval with $c < 0$ and $d > 0$, the density $p(a)$ is fixed (i.e. θ is empty) and $E_a(a) = 0$. Suppose that $A = f(a, \chi) = \chi_2 a + \chi_1$, so that $E(A) = \chi_1$. The domain of A is the interval $\Sigma_\chi = (\chi_2 c + \chi_1, \chi_2 d + \chi_1)$ which is "scaled" by χ_2 and translated by χ_1. Now, if $x = x(A)$ is the dependent function of interest, $E(x)$ is not dependent on χ_2 iff by (5.5)

$$\mathrm{Cov}(\frac{\partial x}{\partial A}, a) = 0$$

Because of the form of f, this condition is equivalent to

$$\mathrm{Cov}\ (\frac{\partial x}{\partial A}, A) = 0$$

An aggregation scheme which obeys this restriction is the linear aggregation approach of Theil, which posits that

$$x = b \cdot A$$

where b is a random variable, distributed with mean β and uncorrelated with A. An extreme case of this condition is where $b(A) = \beta$ with probability 1, which is the LFF case. By (5.4), $E(x)$ is not dependent on χ_1 iff

$$E_A(\frac{\partial x}{\partial A}) = 0$$

which, in the case of Thiel's model requires that $\beta = 0$.

Finally, we note that for this case, higher order derivatives of the aggregate function with respect to χ_1 are just expectations of the same order derivatives of x with respect to A. In particular, the aggregate function is linear in χ_1 if and only if

$$E_A\left(\frac{\partial^2 x}{\partial A^2}\right) = 0$$

which is, of course, satisfied by Thiel's model.

As this example shows, under our simplified changing domain model, some useful results concerning simple domain parameters are possible.[32] The general case of domain change is much more complex, and beyond the scope of this paper.

6. Conclusion

Throughout this paper we have studied the implications on the aggregation process of several common assumptions applied to macroeconomic functions. We began by studying linearity of an aggregate function in a distribution parameter, and found a range of functional form and distribution movement structures implied, containing the notable cases of linear functional form (LFF) and linear shifting behavior of the population distribution (LPM).

We next addressed the assumption of omitting predictors from an aggregate function formulation, where the predictors were parameters of the underlying population distribution. This discussion followed two directions, beginning with the case where the form of the population density was known. In this case covariance restrictions were developed which can be used to statistically confirm or reject parameter omission assumptions.

We next analyzed the assumption that the aggregate function depended only on a certain distribution parameter, with the form of the density unspecified. For discrete distributions, this analysis was made possible by orthogonally decomposing the movement of the density vector into directions and relating the parameter of interest to particular directions.

This development was then applied to the case where the aggregate function depended only on predictor variable mean, again revealing a range of possible functional form and distribution movement structures. LFF assumptions appear as the only case guaranteeing omission of all distribution parameters other than the mean for all possible distributions. Linearity of the aggregate function in the mean showed the relation of LPM structure to the decomposition of the density, as well to LFF structures. Finally, the case of several predictor variable means was seen to generalize the case of a single mean only slightly.

Decomposing the joint distribution movement of two predictor variables reveals the structure under which the aggregate function depends only on the marginal distributions of the predictor variables. LFF assumptions again play an important role, here being equivalent to a behavioral function additive in the two predictor variables.

The final topic of omitting distributional parameters we considered was the case when the population distribution moves in a segmented fashion. Here we found that the segmentation can be ignored only when the aggregate function restricted to each segment was linear and identical across segments. The general analysis of this problem was supplemented by application of the previous decomposition theorems for discrete densities.

Finally, a simple model was presented which analyzed the case of a changing domain of the underlying predictor variable. This model was used to study translating and scaling of the predictor variable domain, and yields conditions akin to the consistent aggregation scheme of H. Theil.

The benefits of this work are in three forms, the first theoretical and the others applicable to empirical work. Theoretically, this work provides the proper focus for unifying previous theoretical work in aggregation theory, as the issues of linear aggregation in its many forms as well as particular aggregation schemes (where distribution and functional forms are assumed and the integration performed directly) are all subsumed in the current framework. The inherent nature of the aggregation problem is the omission of individual and distribution detail in macro relationships, which is exactly our current focus.

The second benefit of this research is in the interpretation of estimated macro functions, as they connect to individual behavioral relations.

As is known from previous work, estimates of macro function parameters will identify the parameters of the individual behavioral relations for arbitrary distribution changes if and only if LFF assumptions obtain, namely the behavioral function is linear (and common across agents) as is the macro function. Here we have presented alternative assumptions guaranteeing a linear macro macro function (notably LPM) which can be empirically checked as to their validity, and afford different macro function interpretations.

An example best illustrates this point. Suppose that A represents income, $x(A)$ consumption, with $E(A)$, $E(x(A))$ average income and average consumption respectively. Suppose that the following linear macro function is estimated and found in accordance with the data:

$$E(x(A)) = a + b\, E(A)$$

In interpreting the coefficient b, authors tend to relate b to the marginal propensity to consume for individual agents. This is only strictly true under LFF assumptions. b is the average marginal propensity to consume only under consistent aggregation (Thiel) assumptions. Alternatively, suppose that the income distribution obeys LPM assumptions with respect to another variable t, say a time trend, as in

$$p(A|t) = p_1(A) + tp_2(A)$$

Then, a true "structural" model for $E(x(A))$ and $E(A)$ can be found as

$$E(x(A)) = c + dt$$

$$E(A) = e + ft$$

where $c = \int x(A)p_1(A)dA$; $d = \int x(A)p_2(A)dA$, $e = \int Ap_1(A)dA$; $f = \int Ap_2(A)dA$

are all constant parameters. The original linear model is just a "reduced form" from this system, with $a = c - \frac{de}{f}$, $b = \frac{d}{f}$; and reveals little about the behavioral function $x(A)$.

This development shows that a serious interpretation of an estimated macro function requires investigation of its aggregation underpinnings. Although there remains much research to be done on the implimentation of an LPM model such as that above, we can at least see the need for some interpretive work, such as the testing of an LFF model on a cross section data base for this example.

The third, and perhaps major, benefit of this research lies in the construction of macroeconomic models. We suppose first that the not so small task of choosing predictor variables is complete, likely through appeal to economic theory. If a macro forecasting model is desired, then simple versions might be justified by studying the movement of the underlying predictor variable distribution. In particular, if LPM assumptions hold, and are expected to continue to hold, then a simple linear forecasting model is justified, without concern for underlying behavioral functional forms.

If a model describing behavior is desired, then our results provide a number of testable sets of assumptions which justify a simplified macro function from one depending explicitly on the entire distribution of predictor variables. Consider first the case where certain of the predictor variables are discrete in nature, so that the population may be segmented by these variables. In view of sections 4.2 and 4.4, a natural starting point is to assume a linear model in each segment between the dependent variables and the remaining (continuous) predictor variables. These linear models can be estimated by standard methods, and linearity tested using the techniques of Stoker (1980). For segments where linear models are justified, tests indicating equality of coefficients can be performed, and if equality is confirmed, these segments can be considered as one in the aggregate function. Completion of such a process will suggest an aggregate formulation which depends on a minimum amount of distribution detail, and which is justifiable via aggregation theory.

The above process illustrates use of the results on segmentation, which are naturally suggested by the discrete predictor variables. If instead, it is desirable to refine a function aggregated over a joint distribution to one depending only on marginal distributions, then additive structures should be sought and tested using micro data.

Although future research must be directed to analyzing the statistical details of the above procedures, the current work serves to indicate the proper structure, either functional form or distribution form, to utilize and test. In each case, successful investigation of aggregation assumptions will permit the simplest aggregate function formulation to be used, which does not omit any necessary distribution detail.

In closing, the overall theme of this paper is that for relations between average variables to be fully understood, one must study both the true behavioral relation for individuals and the movement of the population distribution. Assumptions of the per capita form, namely that average data can be connected by models based only on individual behavior, ignore the true structure of the aggregation process, and can lead to omission of relevant distribution parameters and incorrect interpretation of estimated relations. The importance of the aggregation problem should be underscored, and not ignored, as it often is in current practice.

Appendix: Omitted Proofs

Proof of Theorem 1: Existence of a minimal decomposition is shown by construction: Suppose that there is an orthogonal decomposition of $\underline{p}(\theta)$ of the form

$$\underline{p}(\theta) = \frac{1}{N}\underline{i} + \sum_{i=1}^{N_1} f_i(\theta_0)\underline{\delta}_i + \sum_{j=1}^{N_2} g_j(\theta_0,\theta_1)\underline{\xi}_j \tag{A.1}$$

where $N_1 + N_2 \leq N - 1$, and there exists α_j, $j = 1, .\ .\ N_2$ not all zero, such that $\Sigma\alpha_j g_j(\theta_0,\theta_1) = f(\theta_0)$. Define $\hat{\underline{\delta}} = \Sigma\alpha_j\underline{\xi}_j$ and pick $\hat{\underline{\xi}}_j$, $j = 2, \ldots, N_2$ such that $\left\{\hat{\underline{\delta}}, \hat{\underline{\xi}}_2, \ldots, \hat{\underline{\xi}}_{N_2}\right\}$ is an orthogonal set of vectors, and

$$\text{SPAN}\left\{\hat{\underline{\delta}}, \hat{\underline{\xi}}_2, \ldots, \hat{\underline{\xi}}_{N_2}\right\} = \text{SPAN}\left\{\underline{\xi}_1, \ldots, \underline{\xi}_{N_2}\right\}$$

Transforming the coefficients of $\underline{p}(\theta)$ above to those with respect to $\underline{i}$, $\underline{\delta}_1, \ldots, \underline{\delta}_{N_1}, \hat{\underline{\delta}}, \hat{\underline{\xi}}_2, \ldots, \hat{\underline{\xi}}_{N_2}$ gives

$$\underline{p}(\theta) = \frac{1}{N}\underline{i} + \sum_{i=1}^{N_1} f_i(\theta_0)\underline{\delta}_i + f(\theta_0)\hat{\underline{\delta}} + \sum_{j=2}^{N_2} \hat{g}(\theta_0,\theta_1)\hat{\underline{\xi}}_j$$

Either this decomposition is minimal in θ_1, or the above process can be repeated (less than N-2 times) until a minimal decomposition is found.

For the uniqueness of the subspace $\text{SPAN}\{\underline{\xi}_i\}$, begin by assuming without loss of generality that $\underline{\delta}_1, \ldots, \underline{\delta}_{N_1}, \underline{\xi}_1, \ldots, \underline{\xi}_{N_2}$ and $\underline{\delta}_1^*, \ldots, \underline{\delta}_{N_1}^*, \underline{\xi}_1^*, \ldots, \underline{\xi}_{N_2}^*$ are vectors with norm 1 and that $N_1 + N_2 = N_1^* + N_2^*$. Define the following matrices with these vectors as columns as

$$\Delta = \left[\underline{\delta}_1, \ldots, \underline{\delta}_{N_1}, \underline{\xi}_1, \ldots, \underline{\xi}_{N_2}\right]$$

$$\Delta^* = \left[\underline{\delta}_1^*, \ldots, \underline{\delta}_{N_1^*}^*, \underline{\xi}_1^*, \ldots, \underline{\xi}_{N_2^*}^*\right]$$

Δ and Δ^* are orthogonal matrices, such that $\Delta^{-1} = \Delta'$ and $(\Delta^*)^{-1} = (\Delta^*)'$.
Now define the following vectors of decomposition coefficients

$$F(\theta_0,\theta_1) = \begin{bmatrix} f_1(\theta_0) \\ f_{N_1}(\theta_0) \\ g_1(\theta_0,\theta_1) \\ g_{N_2}(\theta_0,\theta_1) \end{bmatrix}$$

$$F^*(\theta_0,\theta_1) = \begin{bmatrix} f_1^*(\theta_0) \\ f_{N_1^*}^*(\theta_0) \\ g_1^*(\theta_0,\theta_1) \\ g_{N_2^*}^*(\theta_0,\theta_1) \end{bmatrix}$$

By definition, we have

$$\Delta F(\theta_0,\theta_1) = \Delta^* F^*(\theta_0,\theta_1)$$

so that

$$F(\theta_0,\theta_1) = RF^*(\theta_0,\theta_1) \tag{A.2}$$

where $R = \Delta^{-1}\Delta^* = \Delta'\Delta^*$. R is the matrix which changes coefficients with respect to the basis $\left\{\underline{\delta}_i^*,\underline{\xi}_j^*\right\}$ to those of the basis $\left\{\underline{\delta}_i,\underline{\xi}_j\right\}$, and is an orthogonal matrix, since $R^{-1} = (\Delta^*)^{-1}(\Delta')^{-1} = (\Delta^*)'\Delta = R'$. The proof is completed by examining the structure of R: (A.2) is rewritten as

$$\begin{bmatrix} f_1(\theta_o) \\ \vdots \\ f_{N_1}(\theta_o) \\ g_1(\theta_o,\theta_1) \\ \vdots \\ g_{N_2}(\theta_o,\theta_1) \end{bmatrix} = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] \begin{bmatrix} f_1^*(\theta_o) \\ \vdots \\ f_{N_1}^*(\theta_o) \\ g_1^*(\theta_o,\theta_1) \\ \vdots \\ g_{N_2}^*(\theta_o,\theta_1) \end{bmatrix}$$

where A is $N_1 \times N_1^*$, B is $N_1 \times N_2^*$, C is $N_2 \times N_1^*$ and D is $N_2 \times N_2^*$. Now B is a zero matrix, for otherwise there exists α_j, $j = 1, \ldots N_2^*$ such that $\Sigma\alpha_j\, g_j^*(\theta_o, \theta_1) = f(\theta_0)$. Also $N_2 \geq N_2^*$, for otherwise R is singular. Now, noting that

$$F^*(\theta_0,\theta_1) = R'F(\theta_0,\theta_1)$$

and repeating the above argument, we can show that C is a zero matrix, and $N_2^* \geqslant N_2$. Thus $N_2 = N_2^*$, and $\text{SPAN}\{\underline{\xi}_j\} = \text{SPAN}\{\underline{\xi}_j^*\}$, by the block diagonality of R.

Q.E.D.

Proof of Theorem 2:

Suppose that a minimal decomposition of $\underline{p}$ (θ) is denoted as in (A.1). $\underline{x}$ is represented in that basis as

$$\underline{x} = \bar{x}\underline{i} + \sum_{i=1}^{N_1} b_i \, \underline{\delta}_i + \sum_{j=1}^{N_2} c_j \, \underline{\xi}_j + \underline{\varepsilon}$$

where $\bar{x} = \sum_{i=1}^{N} x\,(A_i)/N$, and $\underline{\varepsilon}'\underline{i} = \underline{\varepsilon}'\underline{\delta}_i = \underline{\varepsilon}'\,\underline{\xi}_j = 0$ for all i, j ($\underline{\varepsilon}$ may be the zero vector). E(x) is then

$$E(x) = \underline{x}'\,\underline{p}(\theta) = \bar{x} + \sum_{i=1}^{N_1} b_i \, f_i \, (\theta_o) \, \underline{\delta}_i' \, \underline{\delta}_i$$

$$+ \sum_{j=1}^{N_2} c_j \, g_j \, (\theta_o, \theta_1) \, \underline{\xi}_j' \, \underline{\xi}_j$$

If E(x) is a function ϕ_o (θ_o) of θ_o only, then

$$\phi_o(\theta_o) - \sum_{i=1}^{N_1} b_i \, f_i \, (\theta_o) \, \underline{\delta}_i' \, \underline{\delta}_i$$

$$\equiv f(\theta_0) = \sum_{j=1}^{N_2} c_j \, (\underline{\xi}_j' \, \underline{\xi}_j) \, g_j \, (\theta_o, \theta_1)$$

Since we began with a minimal decomposition, this implies $c_j \, \underline{\xi}_j' \, \underline{\xi}_j = 0$ for all j, or that $c_j = 0$ for all j. Thus $\underline{x} \varepsilon S_{\theta_1}^{\perp}$ Q.E.D.

Theorem 3 is obvious since the function $\theta_1' = g(\theta_1)$ serves only to reparameterize the coefficients g_j (θ_o, θ_1) of a minimal decomposition.

Proof of Theorem 6:

THis result is trivial by applying the technique for proving Theorem 1 twice. First, find a minimal decomposition of $\underline{p}(\theta)$ with respect to $\underline{\theta}_1$, and then repeat the technique, isolating coefficients nonlinear in θ_o. Following this, define R analogously as the matrix transforming two decompositions with the property of Theorem 6, and show as in Theorem 1 that R is block diagonal (with three blocks), proving uniqueness of the subspaces S_η and S_{θ_1}. Proof of the final statement follows the proof of Theorem 2 exactly. Q.E.D.

FOOTNOTES

1. In this paper, models used to explain averaged variables are studied. The terms "aggregate" and "macro" are used interchangeably to describe such models, which include the usual macroeconomic equations such as the consumption and investment functions, as well as so-called "microeconomic" models used to study average production and demand.

2. The consumption-income example is just used for illustration of the basic issues, as opposed to being a special area where aggregation problems are more significant than in other areas. In fact, the classic works on the consumption function, namely Friedman (1957) and Ando and Modigliani (1963) each recognized the distributional foundation.

3. This critique, as well as all other issues addressed in the paper, applies equally well to models explaining total variables (e.g. GNP, total personal consumption expenditures) as to those explaining average variables.

4. There do exist microsimulation models whose foundations are separate behavioral equations for each individual in the population, which are used to forecast average variables by simulating the behavior of each individual, and averaging the results. While such models do attempt to use a tremendous amount of distribution detail, they are not notable for ease of implementation, or economy in time or cost in producing forecasts.

5. See Lau (1980), as well as Gorman (1953), and Muellbauer (1975,1977) for exact aggregation results, which require a common linear model for all individuals. Theil (1959) considers linear models with random coefficients, obtaining aggregation properties through distribution assumptions that coefficient-independent variable covariances vanish. These models are termed "consistent" aggregation models by Barten (1977).

6. This type of assumption was discussed in early studies of the aggregation problem -- see de Woolff (1941) and Farrell (1954). Similar but more elaborate models were pioneered by Houthakker (1956), where both a functional form and distribution form were assumed, and the macro function found by direct integration. More recent work in this area is found in Sato (1975) and MacDonald and Lawrence (1978).

7. Actually this statement is a bit unfair to the works in exact aggregation (see footnote 5), where the common linear structure arises as a necessary and sufficient structure for a particular type of aggregate function, when the change in the underlying configuration of individual attributes is arbitrary. Aside from this, the aggregation literature referred to are the areas cited in Notes 5 and 6. Another line of work in aggregation theory, but not addressed here, are papers dealing with statistical issues of implementing linear models with random coefficients; see, for example, Kuh (1959), Griliches and Grunfeld (1960) and Kuh and Welsh (1976).

8. For example, in Section 4.3, where we consider macro functions depending on marginal distributions where aggregation occurs over a joint distribution.

9. We refer here to situations justifying the Weak Law of Large Numbers (c.f. Rao (1973)). This requires a suitably large population (which exists in most states and countries) as well as the existence of the mean and variance of $x(A)$, which we assume for general distributions and is true for discrete distributions.

10. These parameters would consequently appear as arguments of $E(x) = \phi(\theta)$.

11. All integration (as in (2.1)) is performed over this constant domain, which is assumed to be a measurable subset of the real line R. $x(A)$ is a function from this domain to the real line.

12. LFF is the dominant form of assumption appearing in the exact aggregation literature (c.f. Note 5).

13. θ may be any distributional parameter here, not just the mean of A.

14. Note that θ is constrained to lie in an interval $\Theta = [\theta',\theta'']$ by the condition that $p(A|\theta) \geq 0$ for all A. For $p_1(A)$ to be a true base density, we must have $p_1(A) \geq 0$ for all A. If this is not true, then we can pick $\theta^* \varepsilon \Theta$ and redefine (3.2) as

$$p_1(A) + \theta p_2(A) = (p_1(A) + \theta^* p_2(A)) + (\theta-\theta^*)p_2 A = p_1{}^*(A) + (\theta-\theta^*)p_2(A)$$

where now $p_1{}^*(A)$ is a true base density.

15. Again θ and $d(\theta)$ must be constrained to insure that $p(A|\theta) \geq 0$ for all A.

16. Assumption 2 is unnecessary for the analysis of discrete distributions to follow.

17. See Ferguson (1967) for a statement of this form, as well as several properties of distributions of the exponential family.

18. For details on the reparameterization as well as a derivation of this result, see Stoker (1980).

19. The use of cross-section data to investigate aggregate function structure is the topic of Stoker (1980). In particular, the use of regression coefficients (as in Example 2) is justified for a wide class of models obeying "asymptotic sufficiency" conditions, which include LFF models and models with distribution forms admitting sufficient statistics (such as members of the exponential family).

20. The assumption of a constant domain of A implies that $\underline{A}$ and $\underline{x}$ do not vary with θ. Also, we assume that $A_i \neq A_j$ if $i \neq j$, for otherwise we could redefine $p(A|\theta)$ as the marginal distribution of A, without loss of generality.

21. R^N denotes N dimensional Euclidean space. In the text we make heavy use of the usual vector space and inner product structure of R^N. Textbooks covering the relevant mathematics are Curtis (1968), Hadley (1961) and Gantmacher (1960).

22. Another interpretation is that $p(A|\theta)$ is a linear combination of orthogonal variables whose means are their coefficients; i.e. if in (4.3) $\underline{\delta}_i{}'\underline{\delta}_i = 1$, then $E(\delta_i) = \underline{\delta}_i{}'\underline{p}(\theta) = d_i(\theta)$. Note that the interpretation also applies to (4.2)

23. The notation "SPAN" is defined as

$$\text{SPAN}\left\{ \underline{\xi}_1,\ldots,\underline{\xi}_{N_2} \right\} = \left\{ \sum_{i=1}^{N_2} \alpha_i \underline{\xi}_i \;\middle|\; \alpha_i \in R \right\}$$

i.e. the subspace of R^N generated by the vectors $\underline{\xi}_1,\ldots,\underline{\xi}_{N_2}$.

24. If $\underline{p}(\theta^*)$ is parameterized by an arbitrary vector θ^*, then we can reparameterize $\underline{p}$ by $\theta = (\mu,\theta_1)$ if μ is a nontrivial function of one component of θ^*, through $\mu = \underline{A}'\underline{p}(\theta^*)$. Otherwise μ is constant for all possible values of θ^*.

25. This decomposition, as many in the examples and further formulae in the paper uses the Gram-Schmidt orthogonalization process (c.f. Curtis (1968) or Gantmacher (1960)). In this case $b = \underline{x}'\underline{\tilde{A}}/\underline{\tilde{A}}'\underline{\tilde{A}}$ and $\varepsilon = \underline{x} - \bar{x}\underline{i} - b\underline{\tilde{A}}$.

26. Recall that d_i, $i=3,\ldots,N$ are constrained so that $p(A|\mu,d_3,\ldots,d_N) \geq 0$ for all A. This requires only inequality constraints, since the unit simplex T is N-1 dimensional, and therefore adds no complications.

27. "Δ" refers to differencing; i.e. if $\underline{p}$ changes from $\underline{p}(\mu,Q,R)$ to $\underline{p}(\mu',Q',R')$ then $\Delta Q = Q'-Q$, $\Delta R = R'-R$.

28. Again we apply the Gram-Schmidt process - c.f. note 25.

29. Of course, if specific forms of $x(A,B)$ and $p(A,B|\theta)$ are assumed with $E(x)$ found by integration, then independence of A and B provides tremendous computational advantages.

30. The reader may wonder why this problem is not addressed as one of marginal distributions by defining D = 1 for segment 1 and D = 0 for segment 2, and then considering the joint distribution of A and D. The reason for this is that if $M \neq N$, certain cells of the joint distribution have zero

(con'd) 30. probability. Incorporation of these restrictions in the joint distribution is quite messy, and detracts from illustrating the decomposition structure for conditional distributions.

31. See Aczel (1966).

32. Actually, the original exact aggregation problem of Gorman (1953) (and the other citations of note 5) can most easily be put in the form of the simple model presented here. Let $a \varepsilon \{1,\ldots,N\}$ and $p(a=i) = \frac{1}{N}$. Let $A_i = f(i, \chi) = \chi_i$ (χ is an N vector), so that the mean of A is $\mu = \Sigma A_i/N = \Sigma \chi_i/N$. E(x) is then

$$E(x) = \frac{\Sigma x(\chi_i)}{N} = \frac{\sum_{i=1}^{N-1} x(\chi_i)}{N} + \frac{x(N\mu - \sum_{i=1}^{N-1} \chi_i)}{N}$$

$$= \phi(\mu, \chi_1, \ldots, \chi_{N-1})$$

By differentiating, we see that $\chi_1,\ldots,\chi_{N-1}$ can be omitted from E(x) iff $\frac{\partial x}{\partial A}$ is constant. Integrating gives precisely the exact aggregation result, that x(A) obeys LFF conditions.

REFERENCES

Aczel, J., Lectures on Functional Equations and Their Applications, Academic Press, New York, 1966.

Ando, A. and F. Modigliani, "The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Tests," American Economic Review, March 1963.

Barten, A. P., "The Systems of Consumer Demand Functions Approach: A Review," Econometrica 45 (January 1977):23-52.

Curtis, C. W., Linear Algebra, An Introductory Approach, 2nd ed., Allyn and Bacon, Inc., Boston, 1968.

Farrell, M. J., "Some Aggregation Problems in Demand Analysis," Review of Economic Studies 21 (1954):193-203.

Ferguson, T. S., Mathematical Statistics, A Decision Theoretic Approach, New York: Academic Press, 1967.

Friedman, M., A Theory of the Consumption Function, NBER No. 63, general series, Princeton, New Jersey: Princeton University Press, 1957.

Gantmacher, F. R., The Theory of Matrices, Vol. I, Chelsea, New York, 1960.

Gorman, W. M., "Commodity Preference Fields," Econometrica 21 (January 1953):63-80.

Griliches, A. and Y. Grunfeld, "Is Aggregation Necessarily Bad?," Review of Economics and Satistics, 42, 1960, pp. 1-13.

Hadley, G., Linear Algebra, Addison-Wesley, Reading, Massachusetts, 1961.

Houthakker, H. S., "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis," Review of Economic Studies 23 (1956):27-31.

Kuh, E., "The Validity of Cross-Sectionally Estimated Behavior Equations in Time Series Applications," Econometrica, April 1959.

Kuh, E. and R. Welsh, "The Variances of Regression Coefficient Estimates Using Aggregate Data," Econometrica, February 1976.

Lau, L. J., "Existence Conditions for Aggregate Demand Functions," forthcoming in Econometrica, 1980.

MacDonald, R. J. and A. G. Lawrence, "Aggregate CES Input Demand with Polytomous Micro Demand," Econometrica 46 (1978), pp. 365-78.

Muellbauer, J., "Aggregation, Income Distribution and Consumer Demand," Review of Economic Studies 42 (October 1975):525-43.

Muellbauer, J., "Commodity Preferences and the Representative Consumer," Econometrica 44 (1977):979-99.

Rao, C. R., Linear Statistical Inference and Its Applications, 2nd ed., New York: Wiley (1973).

Sato, K., Production Functions and Aggregation, Amsterdam: North Holland (1975).

Stoker, T. M., "Statistical Aggregation Analysis: Characterizing Macro Functions with Cross Section Data," Working Paper, Sloan School of Management, MIT, April 1980.

Theil, H., Linear Aggregation of Economic Relations, Amsterdam: North Holland, 1954.

Wolff, P. de., "Income Elasticity of Demand, A Micro-Economic and a Macro-Economic Interpretation," Economic Journal 51 (April 1941):104-45.